

**Searches for Lepton Flavour Violating
Decays of the Higgs Boson in pp Collisions
with the ATLAS Detector**

**利用ATLAS探測器搜索在質子對撞中的
希格斯玻色子的輕子味數守恆破缺衰變**

WU, Minlin
吳敏麟

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Physics

The Chinese University of Hong Kong

September 2023

Thesis Assessment Committee

Professor CHU Ming Chung (Chair)

Professor Luis Roberto FLORES CASTILLO (Thesis Supervisor)

Professor NG Chun Yu (Committee Member)

Professor FERNANDEZ TELLEZ Arturo (External Examiner)

Abstract

The Standard Model (SM) of particle physics accurately classifies elementary particles and describes their interactions based on experimental observations and theoretical predictions. However, it lacks explanations for certain phenomena like neutrino oscillations. Motivated by the search for a complete theory of nature, physicists aim to extend the SM by exploring experimental setups which may provide indications of new physics. One of the main goals of the Large Hadron Collider (LHC) experiments is to search for signals that violate the SM predictions, including the signal of the lepton-flavour-violating (LFV) decays of the Higgs boson.

This thesis presents two direct searches for lepton flavour violation in Higgs boson decays, $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$, using data from the ATLAS detector at the LHC. The analysis is based on the dataset of proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1} .

Two background estimation techniques, the MC-template method and the symmetry method, are utilized in this analysis. The MC-template method relies on data-corrected simulation samples, while the symmetry method exploits the symmetry between electrons and muons in the SM backgrounds.

This thesis focuses on the studies of the hadronic decays ($\tau \rightarrow \text{hadrons } \nu_\tau$) of the τ -lepton using the MC-template method. The symmetry method is also briefly described. Additionally, the leptonic decay ($\tau \rightarrow \ell \nu_\ell \nu_\tau$) channels and the combined results are described.

No significant excess of events was observed in the searches for the lepton-flavour-violating decays of the Higgs boson. The results are interpreted as upper limits on the branching ratios, $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$, determined at 95% confidence level. The observed (expected) branching ratio limits are $\mathcal{B}(H \rightarrow e\tau) < 20\%$ (0.12%) and $\mathcal{B}(H \rightarrow \mu\tau) < 0.18\%$ (0.09%), obtained using the MC-template method. The results obtained using the symmetry method are represented as the branching ratio difference, and the best-fit value of $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$, measured in the leptonic decay channels, is $(0.25 \pm 0.10)\%$, which is compatible with zero within a significance of 2.5σ .

The prospect of the searches of LFV decays of the Higgs boson with 3000 fb^{-1} data samples of proton-proton collisions at $\sqrt{s} = 14$ TeV using the ATLAS detector at the High-luminosity LHC (HL-LHC) is presented. The expected HL-LHC results are estimated by extrapolating the dataset with integrated luminosity of 138 fb^{-1} . In comparison, the expected upper limits obtained using the HL-LHC data for the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ are more stringent by a factor of 4.8 and 3.9, respectively.

摘要

粒子物理的標準模型根據實驗觀測和理論預測，準確地對基本粒子進行分類以及描述其相互作用。然而，它無法解釋一些現象，比如中微子振蕩。在尋求完整自然理論的驅動下，物理學家致力於通過探索新物理來擴展標準模型。大型強子對撞機（LHC）實驗的主要目標之一是搜索違反標準模型預測的信號，包括呈現與希格斯玻色子的輕子味數守恆破缺衰變的信號。

本論文介紹了使用LHC上ATLAS探測器的數據對希格斯玻色子的兩種輕子味守恆破缺衰變， $H \rightarrow e\tau$ 和 $H \rightarrow \mu\tau$ ，進行直接搜尋的研究。分析基於具有中心質能量 $\sqrt{s} = 13$ TeV的質子對撞數據集，相對應的積分亮度為 138 fb^{-1} 。

本分析使用了兩種背景估計技術：MC模板方法和對稱方法。MC模板方法依賴於經過數據修正的模擬樣本，而對稱方法利用了標準模型背景中電子和繆子之間的對稱性。

本論文重點研究了使用MC模板方法對 τ 介子的強子衰變（ $\tau \rightarrow \text{hadrons } \nu_\tau$ ）進行的研究。同時簡要描述了對稱方法。此外，還描述了關於 τ 介子的電子和繆子的輕子衰變通道（ $\tau \rightarrow \ell \nu_\ell \nu_\tau$ ）以及綜合結果。

在對希格斯玻色子的輕子味數守恆破缺衰變搜尋中，並未觀測到有任何顯著信號的事件。研究結果被解釋為95%信心水平下的分支比上綫，即 $\mathcal{B}(H \rightarrow e\tau)$ 和 $\mathcal{B}(H \rightarrow \mu\tau)$ 。使用MC模板方法得到的觀測（預期）分支比值為 $\mathcal{B}(H \rightarrow e\tau) < 20\%$ (0.12%)和 $\mathcal{B}(H \rightarrow \mu\tau) < 0.18\%$ (0.09%)。而對稱方法得到的結果則表示為分支比值的差異，以及在輕子味衰變通道中測量的 $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$ 的最佳擬合值為 $(0.25 \pm 0.010)\%$ ，在 2.5σ 的顯著性內與零相容。

本論文還介紹了在具有 3000 fb^{-1} 質子對撞數據樣本、 $\sqrt{s} = 14$ TeV的高亮度LHC（HL-LHC）上搜索希格斯玻色子的輕子味數守恆破缺衰變的前景。通過對具有 138 fb^{-1} 積分亮度的數據進行外推，估測了預期的HL-LHC結果。相比之下，使用HL-LHC數據得到的 $\mathcal{B}(H \rightarrow e\tau)$ 和 $\mathcal{B}(H \rightarrow \mu\tau)$ 的預期上限更為嚴格，分別提高了4.8倍和3.9倍。

Acknowledgement

I would like to take this opportunity to express my sincere appreciation to all those who have provided support throughout my journey as a physics PhD student at the Chinese University of Hong Kong.

First and foremost, I would like to express my deepest gratitude to my supervisor, professor Luis Roberto FLORES-CASTILLO, for not only accepting me as his student but also for his invaluable expertise, guidance and support throughout my PhD studies. His deep insight in particle physics and extensive research experience have greatly enriched my understanding and knowledge. His sincere concern for students' academic career development and happiness has created a supportive environment where we can fully explore our potential and works toward the goals.

I would like to thank professor Ming-Chung CHU for his efforts in establishing the platform that enabled students like myself to join the ATLAS research, and for his invaluable guidance and support throughout my PhD studies. I also would like to thank professor Tom CHENG for his great support and guidance, I am impressed by his extensive knowledge, the conversations with him always have been enriching and relaxing.

I am deeply grateful to Dr. Julia Mariana ITURBE PONCE, for her invaluable guidance to help me get start with my analysis. She is exceptionally patient and provided me detailed instructions. Her guidance and insights have played a crucial role in making significant progress. I sincerely wish her all the best in the future with her family.

I also would like to thank all the teaching and support staff at the physics department of CUHK. I will miss them and will look back with appreciation for this warm and supportive group.

The analysis presented in this thesis is based on the ATLAS group at CERN, I would like to thank all the individuals who have contributed to the LHC experiments. Especially the members of the Higgs LFV analysis group, for their supports and successful collaboration. Special thanks for Luca FIORINI and Valerie LANG, who are the analysis contacts, for their invaluable guidance and great support. I also would like to express my appreciation to Johannes JUNGGEBURTH for his guidance and expertise in my ATLAS qualification task in the muon software group.

I would like to express my gratitude to all the members in the Hong Kong cluster team with whom I have collaborated, especially Dr. Zhidong ZHANG, Dr. Kwok-Lam CHU, Ms. Pan TONG, Dr. Jiawei WANG, Mr. Kin-Yip FUNG, Dr. Kevin KWOK and Mr. Wangzheng ZHANG, for their friendship.

I would like to highlight that this project is partially granted and supported by RGC AoE/P-404/18.

For the gratitude that I cannot express in words, I want to convey my sincerest appreciation to my dear family. Thank you for your unconditional support and love.

*Sincere thanks to my high-school teacher - Williamson Lam,
for his enlightening guidance in physics*

Contents

1	Introduction	1
2	The Standard Model	3
2.1	Fermions	4
2.2	Gauge Bosons and Gauge Theories	5
2.2.1	Local Gauge Invariance	6
2.2.2	Quantum Electrodynamics	7
2.2.3	Quantum Chromodynamics	7
2.2.4	Weak Interactions and Electroweak Unification	9
2.3	The Higgs Mechanism	11
2.4	Higgs Boson Production at the LHC	14
2.5	Higgs Decays at the LHC	16
2.6	Lepton Flavour Conservation	17
3	Beyond the Standard Model and Lepton Flavour Violation	20
3.1	Challenges for the Standard Model	21
3.1.1	Neutrino Oscillations	23
3.2	Lepton Flavour Violation	24
3.2.1	LFV Decays of the Higgs Boson	25
3.2.2	Prior Higgs LFV Searches	26
3.3	Higgs LFV predictions in BSM Theories	27
3.3.1	Two-Higgs-Doublet Model	30
3.3.2	Supersymmetry	31
3.3.3	Composite Higgs Model	33
3.3.4	Randall-Sundrum Model	34
3.3.5	Seesaw Mechanism	35
4	The LHC and the ATLAS Detector	38
4.1	The Large Hadron Collider	38
4.1.1	Layout	39
4.1.2	Accelerator Chain	41
4.2	The ATLAS Detector	42
4.2.1	Layout	42
4.2.2	Coordinate System	43
4.2.3	Inner Detector	44

4.2.4	Calorimeters	46
4.2.4.1	Electromagnetic Calorimeter	46
4.2.4.2	Hadronic Calorimeter	48
4.2.4.3	Forward Calorimeter	48
4.2.5	Muon Spectrometer	49
4.2.5.1	Precision-tracking Chambers	50
4.2.5.2	Trigger Chambers	51
4.2.5.3	New Small Wheels	53
4.2.6	Magnet System	54
4.2.7	Forward Detector	54
4.2.8	Trigger and Data Acquisition	55
4.2.9	Summary of Second Upgrades	56
5	Experimental Data and Simulation Samples	58
5.1	Data Processing	59
5.2	Monte Carlo Simulation Samples	60
5.2.1	LFV Signal Samples	61
5.2.2	Background Samples	62
5.2.3	Detector Simulation	63
5.2.4	Digitization	64
5.3	Muon Software Validation	65
5.3.1	Muon Simulation Hits	66
5.3.2	CSC and TGC Digitization	68
5.3.3	Muon Reconstruction	70
5.3.4	Validation	71
5.3.5	Performance Monitoring	73
6	Object Reconstruction and Identification	76
6.1	Tracks	76
6.2	Primary Vertex	77
6.3	Electrons	78
6.4	Muons	80
6.5	Hadronic Taus	81
6.5.1	New Algorithms used in Run-3	83
6.6	Jets	84
6.7	Overlap Removal	85
6.8	Missing Transverse Energy	85
7	Analysis Strategies	87
7.1	MC-template Method	88
7.2	Symmetry Method	88
7.3	Cut-based Analysis	90
7.3.1	Mass Variables	90

8	Event Selection and Categorization	93
8.1	Selection of the $\ell\tau_{\ell'}$ Channel	93
8.1.1	Trigger Selection	93
8.1.2	Object Selection	94
8.1.3	Lepton Assignment	95
8.1.4	Baseline Selection	95
8.1.5	Control Regions	96
8.1.6	Categorization	97
8.2	Selection of the $\ell\tau_{had}$ Channel	98
8.2.1	Trigger Selection	98
8.2.2	Object Selection	98
8.2.3	Baseline Selection	99
8.2.4	Categorization	99
9	Efficiency Correction	101
9.1	Event Efficiency	101
9.1.1	Trigger Efficiencies	102
9.1.2	Muon Efficiencies	102
9.1.3	Electron Efficiencies	102
9.2	Efficiency Correction Strategies	103
10	Background Estimation	106
10.1	MC-template Method	106
10.1.1	$\ell\tau_{\ell'}$ Channel	106
10.1.1.1	Misidentified Lepton Estimation	109
10.1.2	$\ell\tau_{had}$ Channel	111
10.1.2.1	Fake Factor Method	113
10.2	Symmetry Method	120
10.2.1	$\ell\tau_{\ell'}$ Channel	121
10.2.1.1	Misidentified Lepton Estimation	123
10.2.2	$\ell\tau_{had}$ Channel	123
11	Multivariate Analysis and Machine Learning	126
11.1	Neural Network Algorithm	127
11.2	Boosted Decision Tree Algorithm	128
11.2.1	Hyperparameters	129
11.2.2	Cross-validation	130
11.2.3	Performance	130
11.3	Training Strategies	131
11.3.1	Input Variables Optimisation	134
12	Statistical and Systematic Uncertainties	141
12.1	Statistical Uncertainties	141
12.2	Systematic Uncertainties	142
12.2.1	Experimental Uncertainties	142
12.2.1.1	Electron Uncertainties	142

12.2.1.2	Muon Uncertainties	143
12.2.1.3	Tau Uncertainties	143
12.2.1.4	Jet Uncertainties	144
12.2.1.5	Missing Transverse Energy Uncertainties	145
12.2.1.6	Luminosity Uncertainty	145
12.2.2	Uncertainties in the Misidentified Background	145
12.2.3	$Z \rightarrow \ell\ell$ Background Normalization	147
12.2.4	Uncertainties in the Efficiency Corrections	148
12.2.5	Theory Uncertainties	149
12.2.5.1	Background predictions	149
12.2.5.2	Higgs Boson Background and Signal Predictions	150
12.3	Uncertainty Preprocessing	151
13	Statistical Analysis	153
13.1	Inference and Probability Distributions	153
13.1.1	Statistical Inference	154
13.1.2	Discrete Probability Distributions	155
13.1.3	Probability Density Functions	156
13.2	Likelihood Function	156
13.2.1	Maximum Likelihood Method	157
13.2.2	Binned Maximum Likelihood Fits	157
13.2.3	Simultaneous Fits	158
13.3	Hypothesis Testing	158
13.3.1	Profile Likelihood	159
13.3.2	Test Statistic	159
13.3.2.1	Test Statistic for Discovery	161
13.3.2.2	Test Statistic for Upper Limits	162
13.3.3	Asimov Dataset	162
13.4	Fits Setup for the Higgs LFV Analysis	163
13.4.1	Independent 1-POI Fits	163
13.4.2	Simultaneous 2-POI Fits	164
14	Run-2 Analysis Results	166
14.1	Independent Searches for $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$	166
14.1.1	I-POI Fits of $H \rightarrow e\tau$	166
14.1.2	I-POI Fits of $H \rightarrow \mu\tau$	172
14.2	Simultaneous Measurement of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ Signal	177
14.3	Measurement of the Branching Ratio Difference	186
15	High-Luminosity LHC	188
15.1	Extrapolation Procedure	188
15.1.1	Statistical Uncertainties	190
15.1.2	Experimental Uncertainties	190
15.1.3	Uncertainties in the Misidentified Background	190
15.1.4	Uncertainties in the Efficiency Corrections	190

15.1.5	Theory Uncertainties	191
15.2	Extrapolation Results	192
15.2.1	Simultaneous 2-POI Fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ Searches using the MC-template Method	192
15.2.1.1	1-POI fit of the $H \rightarrow e\tau$	195
15.2.1.2	1-POI fit of the $H \rightarrow \mu\tau$	195
15.2.2	Independent 1-POI Fits of the $\ell\tau_{\ell'}$ Final States using Symmetry Method . . .	196
16	Conclusion and Outlook	201

Chapter 1

Introduction

To understand the behaviour of our universe, physicists observe natural phenomena and search for relations and patterns. The studies of the fundamental constituents of matter have led to the development of particle physics and particle colliders, enabling the indirect observation of microscopic phenomena.

The Standard Model (SM) of particle physics is a theoretical framework aimed at classifying the elementary particles and describing their interactions. The great success of the SM in predicting many phenomena and its agreement with experimental observations made it the most successful framework in particle physics. In 2012, CERN announced the discovery of the Higgs boson [1, 2] predicted in 1964, which has further validated the SM.

Although the SM represents the current best understanding of particle physics, not all known phenomena are included and explained. Gravity, although it corresponds to the best-known phenomena and it is one of the four fundamental interactions, is not included in the SM. The observation of neutrino oscillations [3, 4] cannot be explained by the SM, indicating its incompleteness. Therefore, searching for physics beyond the SM has become one of the main goals to find the complete theory of nature. To solve these puzzles and to extend the SM, theoretically, many studies of Beyond Standard Model (BSM) theories have been developed. Experimentally, signals of the predictions given by the BSM theories may occur in the observations from high-energy particle colliders.

The Large Hadron Collider (LHC) at CERN is the largest and highest-energy particle collider in the world, the LHC experiments are not only designed to validate the SM but also to probe for physics beyond the SM. Since the Higgs boson was discovered, physicists have been working on measuring the properties of this new particle and searching for new physics in the Higgs sector. One of the searches carried out in the ATLAS experiment, which is also the objective of the analysis described in this thesis, is the search for Lepton Flavour Violation (LFV) in decays of the Higgs boson to pairs of leptons. In the SM, lepton flavour is an exact symmetry, but the observation of neutrino oscillations indicates that LFV processes occur in nature.

This thesis starts by introducing the basic theoretical framework of the Standard Model, and the lepton flavour concepts under the SM are also given in chapter 2. In chapter 3, some existing discrepancies with the observations are introduced. Some of the latest experimental results which contradict the SM predictions are also discussed to provide a brief overview of the recent progress in particle physics. Many BSM theories have been proposed to solve those puzzles in the Standard Model, and many of them have predicted LFV decays of the Higgs boson; some of these theories are introduced in chapter 3.

The experimental infrastructure used to carry out this work - the LHC and the ATLAS detector - is introduced in chapter 4, some major upgrades of the ATLAS detector in Run-3 compared to Run-2 are also mentioned. The algorithms used for data production, data processing and object reconstruction at ATLAS are described in chapter 5 and chapter 6.

In this analysis, the direct searches of the LFV decays of the Higgs boson, $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ are presented. Two independent methods with different background estimation techniques are exploited: the MC-template method and the Symmetry method [5]. The former uses the templates from Monte Carlo (MC) simulation, the latter is based on the premise that the kinematic properties of the SM background are symmetric under exchange of electron e and muon μ . The analysis strategies and the event selection are discussed in chapters 7 and 8, respectively. Chapter 9 and chapter 10 introduce the details of the efficiency correction used in the symmetry method and the background estimation in these two methods, respectively. A machine learning technique is employed in this analysis to enhance the classification power between signals and backgrounds; the details are given in chapter 11. The uncertainties considered and the statistical analyses performed are introduced in chapter 12 and chapter 13, respectively.

This thesis focuses on the analysis of the full Run-2 dataset from the ATLAS detector and the corresponding results are presented in chapter 14, while the prior LFV searches are also discussed in section 3.2.2. Throughout the thesis, potential improvements for this analysis are explored, particularly in anticipation of the increased statistics expected from the Run-3 experiments. The expected improvements over the Run-2 analysis results, obtained by extrapolating the total integrated luminosity of the data samples used from 138 fb^{-1} to 3000 fb^{-1} , are evaluated in the High-Luminosity LHC prospect, which is discussed in chapter 15.

Chapter 2

The Standard Model

In the second half of 20th century, a theoretical framework, now known as the Standard Model of particle physics, was developed to describe all known particles and three of the four known fundamental interactions (it does not include gravity). Quantum field theory provides a mathematical framework for the SM. In the context of this theory, particle interactions are described in terms of quantum fields that pervade space-time [6], and the fundamental forces by which elementary particles interact, are described through the exchange of the respective mediators.

Elementary particles in the SM are classified into two classes according to their spin quantum numbers: particles with half-integral spin are called fermions, and those with integral spin are called bosons. The SM describes the three fundamental interactions as mediated by gauge bosons. Fermions and gauge bosons are introduced in this chapter in section 2.1 and section 2.2, respectively.

The four known interactions in nature are gravitational, electromagnetic, strong and weak. The gravitational interaction between ordinary particles is much too weak and it is not described by the SM. The other three interactions in the SM are successfully described by gauge theories and all require local gauge invariance, which is discussed in section 2.2.1. Classical electrodynamics was established by Maxwell, and its quantum version describes the electromagnetic interaction between electrically charged particles, which is discussed in section 2.2.2. Both Quantum Electrodynamics (QED) and Quantum Chromodynamics (QCD) are quantum field theories; the latter describes the strong interaction and is discussed in section 2.2.3. The weak interaction is described by quantum flavourdynamics; however, this theory is rarely used, as the weak interaction is better understood in terms of electroweak processes, which are described by a unified framework that includes the weak interaction and electromagnetism, this is discussed in section 2.2.4. The Standard Model incorporates the theories of quantum electrodynamics (QED), quantum chromodynamics (QCD), and the electroweak theory to describe the fundamental interactions.

Gauge bosons are force mediators, while the Higgs boson is the only elementary scalar boson in the SM and it is responsible for the mass generation of gauge bosons without breaking the gauge theories. The Higgs mechanism, the production modes and the decay modes of the Higgs boson are introduced from section 2.3 to section 2.5, respectively. The flavour symmetry concepts in the SM are discussed in section 2.6.

2.1 Fermions

There are two families of fermions: quarks and leptons, which are classified by the interactions that they participate in and by the charges they carry. In the SM, there are six types, also known as flavours, of quarks and leptons, and each fermion also has its own antiparticle with the same mass but different charge, which could be electric, color and weak. Elementary fermions are further classified into three generations, each consisting two quarks and two leptons. Particles in the second and third generation are the similar to those in the first generation but with different mass.

One of the distinctive features of fermions is that they must obey the *Pauli Exclusion Principle* and follow *Fermi-Dirac statistics*. In the same quantum system, two identical and indistinguishable fermions cannot occupy the same quantum state simultaneously, at least one of their quantum numbers, such as spin, must be different.

Particles carrying color charge can participate in the strong interactions, and only quarks, among all elementary particles, carry both electric charge and color charge, and hence participate in all three fundamental interactions. There are six flavours of quarks: *up* (u), *down* (d), *charm* (c), *strange* (s), *top* (t) and *bottom* (b). Table 2.1 summarises the mass, electric charge (Q), *isospin* (I_3) and baryon number (B) of quarks, as well as the flavour quantum numbers: the *upness* (U), *downness* (D), *charm* (C), *strangeness* (S), *topness* (T') and *bottomness* (B').

Particle	Mass [MeV/ c^2]	Q	I_3	B	U	D	C	S	T'	B'
First generation										
up	2.16	$+\frac{2}{3}$	$+\frac{1}{2}$	$\frac{1}{3}$	1	0	0	0	0	0
down	4.67	$-\frac{1}{3}$	$-\frac{1}{2}$	$\frac{1}{3}$	0	-1	0	0	0	0
Second generation										
charm	1270	$+\frac{2}{3}$	0	$\frac{1}{3}$	0	0	1	0	0	0
strange	93.4	$-\frac{1}{3}$	0	$\frac{1}{3}$	0	0	0	-1	0	0
Third generation										
top	173000	$+\frac{2}{3}$	0	$\frac{1}{3}$	0	0	0	0	1	0
bottom	4180	$-\frac{1}{3}$	0	$\frac{1}{3}$	0	0	0	0	0	-1

Table 2.1: The properties of quarks [7] in the SM. All anti-quarks have the same mass but with opposite sign of these quantum numbers.

Isolated quarks have not been observed individually due to *color confinement* [8]. However, they can be observed within composite particles, known as hadrons, which consist of two or more quarks bound together by the strong force. Hadrons are subatomic particles consisting of either three quarks (baryons) or a quark and an antiquark (mesons). No individual quark can be observed; instead, they cluster together to form *hadronic jets* which can be detected.

Each lepton generation includes one electrically charged lepton and one electrically neutral, the neutral one also known as a *neutrino* (ν). The six flavours and the properties of leptons are summarised in table 2.2 with their mass, electric charge and *lepton flavor number* (L_e , L_μ , L_τ).

Particle	Symbol	Mass [MeV/ c^2]	Q	L_e	L_μ	L_τ
First generation						
electron	e	0.51	$-\frac{1}{2}$	1	0	0
electron neutrino	ν_e	< 0.0000011	0	1	0	0
Second generation						
muon	μ	106	$-\frac{1}{2}$	0	1	0
muon neutrino	ν_μ	< 0.0000011	0	0	1	0
Third generation						
tau	τ	1777	$-\frac{1}{2}$	0	0	1
tau neutrino	ν_τ	< 0.0000011	0	0	0	1

Table 2.2: The properties of leptons [7] in the SM. All anti-leptons have the same mass but with opposite sign of electric charge and lepton flavor numbers.

The electron, the well-known elementary particle, has the lowest mass of all charged leptons, and the history of particle physics began when the electron was discovered by J. J. Thomson in 1897 [9]. Muons and taus are much heavier and can only be produced in high energy collisions. Whereas neutrinos do not carry electric charge and their masses are so small that they cannot be measured directly with any current technology. In the ATLAS experiment, the presence of the neutrinos is the missing energy.

2.2 Gauge Bosons and Gauge Theories

Elementary bosons in the Standard Model are classified as vector bosons and scalar boson according to their spin. The vector bosons have a spin quantum number of 1, they are described by gauge theories and are also known as gauge bosons. In quantum field theory, gauge bosons are the force carriers of the three fundamental interactions among elementary fermions. The properties of each gauge boson and the respective force are summarised in table 2.3. In the SM, there is only one elementary scalar boson (spin quantum number of 0) which is the Higgs boson. This particle is responsible for the mass of gauge bosons as discussed later in section 2.3.

Two identical and indistinguishable bosons do not follow the Pauli exclusion principle as the corresponding wavefunctions are symmetric under exchange, unlike those of fermions. An example of identical bosons is described by the Bose-Einstein statistics.

Mediator	Symbol	Force	Mass [GeV/ c^2]	Interaction Range [m]	Charge
photon	γ	Electromagnetic	0	Long-range	0
gluon	g	strong	0	10^{-15}	0
Z boson	Z	weak	91.19	10^{-18}	0
W boson	$W^\pm$		80.38		± 1

Table 2.3: The properties of gauge bosons or mediators [7] in the SM.

2.2.1 Local Gauge Invariance

If a system remains unchanged after an operation, it is considered symmetric under that operation; such symmetries can be found in many natural phenomena. An *invariant* is an observable of a system which remains unchanged after certain transformation. Gauge theory is a field theory where the Lagrangian remains invariant under local continuous transformations, and gauge bosons represent the discrete units of gauge fields. Taking the following Lagrangian of a vector field as example:

$$\mathcal{L}_{Proca} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A^\nu A_\nu \quad (2.1)$$

where the tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ and A^μ is a gauge field. Applying the Euler-Lagrange equation to A^ν in equation 2.1 gives the field equation:

$$\partial_\mu F^{\mu\nu} + \left(\frac{mc}{\hbar}\right)^2 A^\nu A_\nu = 0 \quad (2.2)$$

Known as the *Proca equation*, which is a relativistic wave equation which describes a free particle with mass m and spin 1. For a massless field, i.e., for $m = 0$, this corresponds exactly to the Maxwell's equations for empty space.

Similar to the Proca equation, the wave equation of free particles with spin- $\frac{1}{2}$ and with even parity is obtained from the solutions of the *Dirac equation*:

$$i\gamma^\mu \partial_\mu \psi - \left(\frac{mc}{\hbar}\right)\psi = 0 \quad (2.3)$$

which is derived from the *Dirac Lagrangian*:

$$\mathcal{L}_{Dirac} = i\hbar c \bar{\psi} \gamma^\mu \partial_\mu \psi - mc^2 \bar{\psi} \psi \quad (2.4)$$

and it is invariant under global phase transformation:

$$\psi \rightarrow e^{i\theta} \psi \quad (2.5)$$

whereas the free Dirac Lagrangian is not locally phase invariant, and the symmetry is broken under a local phase transformation - when the phase factor in equation 2.5 becomes spacetime-dependent:

$$\mathcal{L}_{Dirac} \rightarrow \mathcal{L}_{Dirac} - \hbar c (\partial_\mu \theta) \bar{\psi} \gamma^\mu \psi \quad (2.6)$$

The extra term that arises in equation 2.6 is the price for preserving the invariance, and what has to be paid is introducing a new vector field A_μ which couples to ψ . The vector field A_μ is a gauge field; the local symmetry can be represented by the gauge transformation with a parameter related to the phase factor $\lambda(x) = -\frac{\hbar c}{q}\theta(x)$:

$$A_\mu \rightarrow A_\mu + \partial_\mu \lambda \quad (2.7)$$

In addition to the coupling term with ψ , the free term of its Lagrangian should be taken into account through the massless version of the Proca Lagrangian or Maxwell's equations given in equation 2.1, as the mass term violates the local invariance. The locally invariant Lagrangian can be written as:

$$\mathcal{L} = [i\hbar c \bar{\psi} \gamma^\mu \partial_\mu \psi - mc^2 \bar{\psi} \psi] - \frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} - (q \bar{\psi} \gamma^\mu \psi) A_\mu \quad (2.8)$$

where q is the coupling constant.

The operation to convert the global symmetry to a local one can be expressed by the *covariant derivative* $\mathcal{D}_\mu \equiv \partial_\mu + i \frac{q}{\hbar c} A_\mu$, which represents the covariant version of a gauge theory:

$$\mathcal{D}_\mu \psi \rightarrow e^{-iq\lambda/\hbar c} \mathcal{D}_\mu \psi \quad (2.9)$$

Hence, the Lagrangian can be simplified as:

$$\mathcal{L} = \bar{\psi} (i \gamma^\mu \mathcal{D}_\mu - m) \psi - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \quad (2.10)$$

2.2.2 Quantum Electrodynamics

In quantum electrodynamics, the gauge field A_μ is exactly the covariant four-potential of the electromagnetic field originating from the electron. Equation 2.8 is the QED Lagrangian involving the Dirac term, the Maxwell term and the interaction term:

$$\mathcal{L}_{QED} = \mathcal{L}_{Dirac} + \mathcal{L}_{EM} + \mathcal{L}_{interaction} \quad (2.11)$$

The changes in the Dirac Lagrangian after demanding local gauge invariance arise in the interaction term. In the context of quantum field theory, the quanta of the gauge field are known as *gauge bosons*. In QED, the photon is the gauge boson or mediator of the electromagnetic interactions, all electrically charged particles interact by exchanging photons.

The gauge invariance under the transformation belongs to the abelian group of the U(1) symmetry in gauge theory. In QED, the Lagrangian admits both global and local U(1) gauge invariance. Noether's theorem states that every conservation law has a corresponding symmetry in nature [10], the U(1) symmetry in QED leads to the conservation of charge, which can be derived from the Maxwell Lagrangian - the last two terms of equation 2.8 with the current density:

$$J^\mu = cq(\bar{\psi} \gamma^\mu \psi) \quad (2.12)$$

2.2.3 Quantum Chromodynamics

An analog of electric charge is the color charge, which is a property of quarks and gluons. There are three types of color which are labelled as *red*, *green* and *blue*, and each of them has a corresponding anticolor - *antired*, *antigreen* and *antiblu*e. The interactions between the color-charged particles are described by quantum chromodynamics. For a certain quark, the free Lagrangian can be written in term of the three colors:

$$\mathcal{L} = [i\hbar c \bar{\psi}_r \gamma^\mu \partial_\mu \psi_r - mc^2 \bar{\psi}_r \psi_r] + [i\hbar c \bar{\psi}_g \gamma^\mu \partial_\mu \psi_g - mc^2 \bar{\psi}_g \psi_g] + [i\hbar c \bar{\psi}_b \gamma^\mu \partial_\mu \psi_b - mc^2 \bar{\psi}_b \psi_b] \quad (2.13)$$

It can be rewritten as the same form of equation 2.4 by representing the four-component Dirac spinors as a triplet $\psi = (\psi_r, \psi_g, \psi_b)^T$. Equation 2.13 also admits global invariance under the transformation similar to equation 2.5 but with a 3×3 Hermitian matrix H as the phase factor:

$$\psi' \rightarrow e^{iH}\psi, \quad H = \theta \mathbf{I} + \boldsymbol{\lambda} \cdot \mathbf{a} \quad (2.14)$$

where $\mathbf{I}$ is a unit matrix, $\boldsymbol{\lambda}$ are the *Gell-Mann matrices* (there are eight of them [11]), and $\mathbf{a}$ are eight real numbers. This global symmetry follows the non-abelian group SU(3).

In 1954, Chen Ning Yang and Robert Mills extended the ideas of abelian groups to non-abelian groups of SU(3) symmetry [12] to explain the strong interactions, which is known as the *Yang-Mills theory*. The QCD Lagrangian which admits the local SU(3) invariance can also be obtained by using the covariant derivative:

$$\mathcal{D}_\mu \equiv \partial_\mu + ig_s \boldsymbol{\lambda}/2 \cdot \boldsymbol{\mathcal{A}}_\mu \quad (2.15)$$

where the coupling is denoted by g_s (analogous to the coupling constant q in the QED), and $\boldsymbol{\mathcal{A}}_\mu$ are the gauge fields introduced by the covariant derivative. There are eight of them which act as an SU(3) octet. The corresponding quanta of the eight gauge fields are known as *gluons*, which mediate strong interactions.

The QCD Lagrangian also demands the free term of the gluon fields, which can be written in terms of the gauge invariant gluon field strength tensor $G_{\mu\nu}^a$:

$$\mathcal{L}_{gluons} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad G_{\mu\nu}^a = \partial_\mu \mathcal{A}_\nu^a - \partial_\nu \mathcal{A}_\mu^a + g_s f^{abc} \mathcal{A}_\mu^b \mathcal{A}_\nu^c \quad (2.16)$$

where $a = 1, 2, \dots, 8$ and f^{abc} are the structure constants of SU(3). The QCD Lagrangian which admits the local SU(3) invariance, is:

$$\mathcal{L}_{QCD} = \bar{\psi}_i [i\gamma^\mu (\mathcal{D}_\mu)_{ij} - m\delta_{ij}] \psi_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (2.17)$$

The gluon fields $\boldsymbol{\mathcal{A}}_\mu$ in equation 2.17 describe the propagation of gluons and interact with the Dirac fields ψ which represent the quarks, similar to the electromagnetic interactions mediated by the photon. However, one distinct feature of gluon fields is that the field lines between two color charges form a narrow flux tube, and the field energy grows as the separation between the color charges increases. The strong force between them is independent of separation, and the energy gained by separating them forms a new quark-antiquark pair instead of further increasing the length of the field lines. This phenomenon is known as *color confinement* [8] as described in section 2.1.

The U(1) symmetry in QED gives the conservation law of electric charge, as given in equation 2.12, according to Noether's theorem. Similarly, the SU(3) flavour symmetry gives the conservation law of hypercharge in QCD. The concept of hypercharge relates electric charge, isospin and flavour into one formula, known as *Gell-Mann-Nishijima formula* [13]:

$$Q = I_3 + \frac{1}{2}Y, \quad Y = B + S + C + B' + T' \quad (2.18)$$

where Y is the hypercharge, and other quantum numbers are introduced in table 2.1. Isospin is conserved in strong interactions and is a subset of flavour symmetry. Hypercharge is a useful quantum number for classifying hadrons, but it is not a conserved quantity in weak interactions.

2.2.4 Weak Interactions and Electroweak Unification

For a system with two spin- $\frac{1}{2}$ fields, ψ_1 and ψ_2 , the Lagrangian in the absence of interactions can be obtained by the sum of two Dirac Lagrangians with a doublet ψ whose components are ψ_1 and ψ_2 :

$$\mathcal{L} = [i\hbar c \bar{\psi}_1 \gamma^\mu \partial_\mu \psi_1 - m_1 c^2 \bar{\psi}_1 \psi_1] + [i\hbar c \bar{\psi}_2 \gamma^\mu \partial_\mu \psi_2 - m_2 c^2 \bar{\psi}_2 \psi_2] \quad (2.19)$$

If the two mass terms are equal, the Lagrangian reduces to the Dirac Lagrangian in equation 2.4 with the doublet ψ . Equation 2.19 admits the global SU(2) invariance under the transformation with a 2×2 Hermitian matrix H :

$$\psi' \rightarrow e^{iH} \psi, \quad H = \theta \mathbf{I} + \boldsymbol{\tau} \cdot \mathbf{a} \quad (2.20)$$

where $\boldsymbol{\tau}$ are the *Pauli matrices* and $\mathbf{a}$ are three real numbers.

In the language of Yang-Mills theory, the Lagrangian that admits the local SU(2) invariance can be derived by using the covariant derivative with the three gauge fields $\mathcal{A}_\mu$ introduced:

$$\mathcal{D}_\mu \equiv \partial_\mu + ig_w \boldsymbol{\tau} / 2 \cdot \mathcal{A}_\mu \quad (2.21)$$

where g_w is the coupling constant. The Lagrangian that admits the local SU(2) gauge invariance can also be written in the same form as equation 2.17 but with field strength tensor $F^{\mu\nu}$:

$$F_i^{\mu\nu} \equiv \partial^\mu \mathcal{A}_i^\nu - \partial^\nu \mathcal{A}_i^\mu - g_w \epsilon_{ijk} \mathcal{A}_j^\mu \mathcal{A}_k^\nu \quad (2.22)$$

The three gauge fields in equation 2.22 are known as the *weak isopin fields*: W_1^μ , W_2^μ and W_3^μ , for which the Lagrangian is invariant under the local SU(2) transformation. The *weak isopin* is a conserved quantum number in all three fundamental interactions. All elementary fermions have non-zero weak isopin, either $+\frac{1}{2}$ or $-\frac{1}{2}$. The local SU(2) symmetry requires gauge bosons with weak isospin $T = 1$ and the associated eigenvalues $T_3 = +1, 0, -1$, which they are referred to the W bosons: W^+ , W^0 and W^- , respectively.

Equation 2.18 can also be applied to the *weak hypercharge* Y_W by replacing the isopin I_3 with weak isopin T_3 . The symmetry associated with Y_W is U(1), and the respective gauge field is known as the weak hypercharge field B . Its field strength tensor can be represented as $B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$, which is in the same form as the electromagnetic field tensor.

Both weak isopin fields and the weak hypercharge field belong to the associated gauge symmetry group. The conservation laws associated with these symmetries are defined by the current J_μ in the interaction term of their Lagrangian. The concept of *parity* is introduced to describe the current of the four fields, as Chien-Shiung Wu discovered *parity violation* in the weak interaction in an experiment in 1956 [14]. This discovery and the chiral representation of the currents hinted the possibility of unifying the weak interaction with electromagnetism.

Chiral symmetry refers to the invariance under the parity inversion, which transfers a phenomenon into its mirror image. A chiral phenomenon is one that behaves differently to its mirrored image. Wu's experiment, based on the beta decay of Cobalt 60, proved that the weak interaction is a chiral gauge interaction, and parity is not an universal symmetry in nature. This discovery led to the introduction of vectorial symmetry, where the same transformation applies to both chiralities, and axial symmetry, where one transformation and its inverse affect left-handed and right-handed

particles, respectively. These symmetries effectively describe the transformations of particles with different handedness, illuminating the subtle intricacies of fundamental interactions.

To build up a unified theory of the weak and electromagnetic interactions, the differences between them needs to be investigated. This study is relatively easier under Feynman's language, where the free term in the Lagrangian determines the propagator and the interaction terms determine the vertex factors. The structural difference of the vertex factors can be studied by defining the current in the interaction terms. The electromagnetic vertex factors are purely vectorial or exhibit vectorial symmetry, on the other hand, the γ^μ is included in the interaction term as shown in equation 2.12. While the weak vertex factors exhibit both vectorial and axial symmetry. In particular, the term $\gamma^\mu(1 - \gamma^5)$ determines the $W^\pm$ coupling.

The left-handed and right-handed projection operators acting on the particle spinors u are:

$$u_L(p) \equiv \frac{1 - \gamma^5}{2} u(p), \quad u_R(p) \equiv \frac{1 + \gamma^5}{2} u(p) \quad (2.23)$$

where the antiparticle spinors have inverse projection operators, meaning when the factors in equation 2.23 act on antiparticle bispinors, they return the opposite handedness than in equation 2.23. Comparing this to the $W^\pm$ coupling term, one can conclude that the $W^\pm$ bosons only couple to the left-handed particles and right-handed antiparticles. Therefore, the doublet ψ in the interaction term should be denoted as ψ_L , which can also be illustrated by an example of beta decay $e^- \rightarrow \nu_e + W^-$:

$$J_\mu^- = \bar{\nu} \gamma_\mu \left(\frac{1 - \gamma^5}{2} \right) e \quad (2.24)$$

the current can be expressed in a different way by rewriting the vertex factors as:

$$\gamma_\mu \left(\frac{1 - \gamma^5}{2} \right) = \left(\frac{1 + \gamma^5}{2} \right) \gamma_\mu \left(\frac{1 - \gamma^5}{2} \right) \quad (2.25)$$

Hence the current $J_\mu = \bar{\nu}_L \gamma_\mu e_L$ is purely vectorial and involves only the left-handed electron and the left-handed neutrino.

The electromagnetic current couples to both left-handed and right-handed particles, and that becomes the only structural difference when compared to the weak current vertex. The electromagnetic current and the weak current can be related by the Gell-Mann-Nishijima formula in equation 2.18; the weak hypercharge current J_μ^Y is:

$$J_\mu^Y = 2J_\mu^{EM} - 2J_\mu^0 \quad (2.26)$$

where J_μ^Y refers to the interaction with B field and J_μ^0 refers to the interaction with the W^3 field. The three weak currents can be combined in a general form:

$$\mathbf{J}_\mu = \frac{1}{2} \bar{\psi}_L \gamma_\mu \boldsymbol{\tau} \psi_L \quad (2.27)$$

where $\tau^\pm$ are linear combinations of Pauli matrices¹. The left-handed doublets ψ represent:

$$\psi_L \rightarrow \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \quad (2.28)$$

¹ τ^0 is the third Pauli matrix τ^3 and $\tau^\pm = \frac{1}{2}(\tau^1 \pm i\tau^2)$

where d' , s' and b' are the Cabibbo–Kobayashi–Maskawa (CKM) rotated fields [7, 15]. The J_μ^0 is found to be a neutral weak current by expanding equation 2.27 in terms of, for example, ν_e and e :

$$J_\mu^0 = \frac{1}{2}\bar{\nu}_L\gamma_\mu\nu_L - \frac{1}{2}\bar{e}_L\gamma_\mu e_L \quad (2.29)$$

That requires a neutral mediator to obey the conservation law of electric charge.

The problem of the structural difference and the neutral gauge boson prediction was solved by Sheldon Glashow, who proposed unified theory [16] of the electromagnetic and weak interactions with a larger symmetry group $SU(2)\times U(1)$, and a neutral Z boson. A similar model, also with $SU(2)\times U(1)$ gauge group, was presented by Abdus Salam and John Clive Ward later in 1964 [17], and predicted the massless photon γ and three massive gauge bosons with a broken symmetry. However, gauge bosons with masses were known to be non-renormalizable and the symmetry breaking which gave the $W^\pm$ and Z bosons masses was inserted manually. The goal became finding the unified theory which can accommodate the massive gauge fields, that needs to exploit *spontaneous symmetry-breaking* and the *Higgs mechanism* discussed in section 2.3.

Gerard't Hooft proved that spontaneously broken gauge symmetries are renormalizable even with massive gauge bosons [18]. The Lagrangian is a perturbation procedure and the fields were treated as the fluctuations of vacuum states; spontaneous symmetry-breaking happens when the vacuum states do not share the symmetry of the Lagrangian.

The prediction of the Z boson was interpreted by Weinberg [19] who noticed that the spontaneous symmetry-breaking produces the Z^0 vector boson and the photon γ by rotating the W^0 and B^0 vector boson plane. The angle is known as the *Weinberg angle* or the *weak mixing angle* θ_W :

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \quad (2.30)$$

The weak mixing angle θ_W depends on the electric charge and the couplings of weak isospin g or weak hypercharge g' :

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \quad \sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \quad (2.31)$$

The electric charge $e = g \sin\theta_W = g' \cos\theta_W$. The weak mixing angle θ_W also introduces a difference in masses of the Z^0 and $W^\pm$ bosons:

$$m_Z = \frac{m_W}{\cos\theta_W} \quad (2.32)$$

In this theory, the way in which the symmetry is broken can also explain the generation of mass of the $W^\pm$ and Z bosons without breaking the local gauge symmetry. This is established through the *Higgs mechanism*.

2.3 The Higgs Mechanism

Local gauge invariance in both electromagnetic and strong interactions demands massless gauge fields as the mass term is not locally invariant. However, experimental measurements show that the $W^\pm$ and Z bosons have relatively large masses. A mechanism for explaining the mass generation of the gauge bosons in the Standard Model must be developed.

Consider a system with two fields, similar to the one in section 2.2.4, to construct the full Lagrangian for electroweak interactions. This time, ψ_1 and ψ_2 can be combined into a scalar complex field:

$$\psi = \psi_1 + i\psi_2 \quad (2.33)$$

To take care of the broken continuous symmetry, one needs to find the minima of the potential in the Lagrangian $\mathcal{L} = \mathcal{T} - \mathcal{U}$ or the non-zero vacuum states. Consider the simplest model with Lagrangian:

$$\mathcal{L} = \partial_\mu \psi^* \partial^\mu \psi - \frac{1}{2} \lambda (\psi^* \psi - \frac{1}{2} \eta^2)^2 \quad (2.34)$$

where λ and η are positive constants. The potential term has a maximum at $\psi = 0$ and the expectation value of ψ is non-zero in vacuum state. The potential term and the minima, which lie on a circle, are shown in figure 2.1. The expectation value in the vacuum state is:

$$\langle 0 | \psi | 0 \rangle = \frac{1}{\sqrt{2}} \eta e^{i\alpha} \quad (2.35)$$

where α is a real number and the radius of the circle of minima can be obtained by setting $\alpha = 0$ and equals $\frac{1}{\sqrt{2}} \eta$, which is the vacuum expectation value (vev) of the Higgs field. The occurrence of the vev signals the spontaneous electroweak symmetry breaking at a scale $\eta = 246$ GeV, which is also known as the electroweak scale.

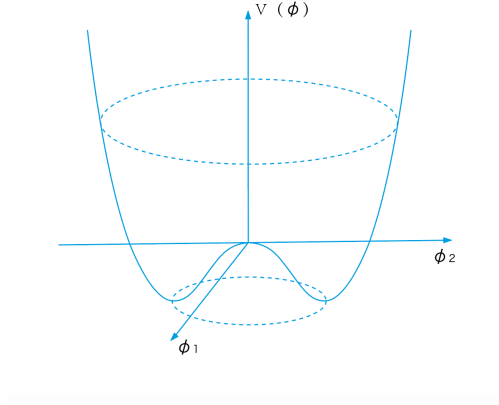


Figure 2.1: The potential function in equation 2.34.

When $\alpha = 0$, the particular vacuum state is:

$$\psi_{1,min} = \frac{1}{\sqrt{2}} \eta, \quad \psi_{2,min} = 0 \quad (2.36)$$

To represent the fluctuations of this vacuum state, two new fields are introduced:

$$\Psi \equiv \psi_1 - \frac{1}{\sqrt{2}} \eta, \quad \Phi \equiv \psi_2 \quad (2.37)$$

Expanding the potential term in equation 2.34 and rewriting it in terms of the new fields gives:

$$\mathcal{L}' = [\frac{1}{2}(\partial_\mu \Psi)(\partial^\mu \Psi) - \lambda \eta^2 \Psi^2] + [\frac{1}{2}(\partial_\mu \Phi)(\partial^\mu \Phi)] - [\sqrt{2}\lambda\eta(\Psi^3 + \Psi\Phi^2) + \frac{1}{2}\lambda(\Psi^4 + \Phi^4 + 2\Psi^2\Phi^2)] \quad (2.38)$$

where irrelevant constants have been removed. The first term in this Lagrangian is a free Klein-Gordon Lagrangian and the third term describes the couplings. The second term introduces a massless field Φ automatically, the appearance of massless bosons is predicted by the *Goldstone's theorem* [20]: massless scalar particles appear when a continuous global symmetry is spontaneously broken. These particles are also known as *Goldstone bosons* which have not been observed.

The massless Goldstone bosons appeared when the global symmetry is spontaneously broken, what about the spontaneously broken *local* symmetry case? Applying the covariant derivative given in equation 2.9 to the Lagrangian:

$$\mathcal{L} = \mathcal{D}_\mu \psi^* \mathcal{D}^\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} \lambda (\psi^* \psi - \frac{1}{2} \eta^2)^2 \quad (2.39)$$

where $F_{\mu\nu}$ is the electromagnetic field strength tensor, and $\mathcal{D}_\mu \psi = \partial_\mu \psi + ieA_\mu \psi$. In the case of $\alpha = 0$, the B field can be represented as:

$$B_\mu = A_\mu - \frac{1}{e\eta} \partial_\mu \psi_2 \quad (2.40)$$

Under gauge transformation, this gives $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. Putting it back into equation 2.39:

$$\mathcal{L} = \frac{1}{2} \partial_\mu \psi_1 \partial^\mu \psi_1 - \frac{1}{2} \lambda \eta^2 \psi_1^2 - [\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{2} e^2 \eta^2 B_\mu B^\mu] + \dots \quad (2.41)$$

The second term gives rise to the Higgs mass $m_H = \eta \sqrt{\frac{\lambda}{2}}$. The second term in the bracket in this Lagrangian is in the same form of the Proca Lagrangian in equation 2.1, which indicates that the massless Goldstone boson have combined to give a massive boson with mass $m \propto g\eta/2$, where g is the $SU(2)$ gauge coupling constant. The rest of the terms in the Lagrangian represent the higher order couplings. This mechanism which without the massless field and only with massive vector fields is known as the *Higgs mechanism*, and the continuous symmetry is broken spontaneously without breaking the gauge symmetries.

In electroweak interactions, the $SU(2) \times U(1)_Y$ symmetry is broken: applying the Pauli matrices σ^i or the hypercharge operator Y on the vacuum state ψ_0 give non-zero results. However, the $U(1)_{EM}$ symmetry is found to be unbroken as $\hat{Q}\psi_0 = 0$, where $\hat{Q} = Y/2 + T_3$ is the electric charge operator. The electroweak symmetry is spontaneously broken from $SU(2) \times U(1)_Y$ to $U(1)_{EM}$.

The Higgs mechanism is triggered by the spontaneous symmetry breaking and, it gives masses to the gauge bosons. The Higgs field is a quantum field that permeates all space. Its interactions with elementary fermions are described by the *Yukawa Lagrangian* term which is denoted as $\mathcal{L}_{Yukawa}$. Hence, the Lagrangian of electroweak theory can be represented as:

$$\mathcal{L}_{EW} = \mathcal{L}_{fermions} + \mathcal{L}_{gauge} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \quad (2.42)$$

This equation can be decomposed into the kinetic term of the elementary fermions:

$$\mathcal{L}_{fermions} = \bar{\psi}_L \gamma^\mu (i\partial_\mu + g \frac{1}{2} \boldsymbol{\tau} \cdot \mathbf{W} + g' \frac{1}{2} Y B_\mu) \psi_L + \bar{\psi}_R \gamma^\mu (i\partial_\mu + g' \frac{1}{2} Y B_\mu) \psi_R \quad (2.43)$$

where ψ_L is the left-handed doublet and ψ_R is the charged lepton singlet. The gauge term describes the interaction between the three W bosons and the B boson:

$$\mathcal{L}_{gauge} = -\frac{1}{4}\mathbf{W}_{\mu\nu}^a \cdot \mathbf{W}_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad (2.44)$$

where $a = 1, 2, 3$. The Higgs term describes the interaction between the Higgs field and the gauge bosons:

$$\mathcal{L}_{Higgs} = |(i\partial_\mu + g\frac{1}{2}\boldsymbol{\tau} \cdot \mathbf{W} + g'\frac{1}{2}YB_\mu)\phi|^2 - [-\frac{1}{2}\lambda\eta^2\phi^*\phi + \frac{1}{2}\lambda(\phi^*\phi)^2] \quad (2.45)$$

After the spontaneous symmetry-breaking, the Higgs doublet $\phi(x)$ is:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.46)$$

where v is the vacuum expectation value and $h(x)$ is the fluctuation around the minimum. Lastly, the Yukawa coupling term:

$$\mathcal{L}_{Yukawa} = -G_-^f(\bar{\phi}_L\phi\phi_R + \bar{\phi}_R\bar{\phi}\phi_L) - G_+^f(\bar{\phi}_L\phi^\dagger\phi_R + \bar{\phi}_R\bar{\phi}^\dagger\phi_L) \quad (2.47)$$

where $G_\pm^f$ are the Yukawa coupling of fermions for $T_3 = \pm\frac{1}{2}$. The masses of the interacting fermions are given by the Higgs field in the state $\psi(x)$ in equation 2.46.

2.4 Higgs Boson Production at the LHC

The quantum excitation of the Higgs field is known as the *Higgs boson*, which is a massive boson carrying zero spin and zero electric charge. Higgs bosons can be produced in the proton-proton collisions at the LHC. The possible production modes of the Higgs bosons are determined through the couplings or the vertices. Equation 2.44 and equation 2.47 describe the couplings of the Higgs bosons to the gauge bosons and fermions. The relative cross sections of the Higgs boson production modes are defined by the coupling strengths.

The Higgs bosons has a relatively large mass, which is measured as 125.25 ± 0.17 GeV/ c^2 [7]. Therefore, the production of the Higgs boson requires sufficient energy to obey the conservation laws. The dominant production modes can be observed from the high-energy particle collisions, benefiting from the current high-energy particle colliders. At the LHC, the dominant contributions come from processes involving quarks and gluons as the collided particles are hadrons (protons). The dominant production modes are:

- (i) Gluon-gluon fusion (ggF): The coupling strength of the particles to the Higgs boson is proportional to their mass. Therefore, processes involving heavy quarks like top and bottom are favoured. The first Feynman diagram in figure 2.2 shows the most likely way to produce a Higgs boson at the LHC: two gluons combine to form a loop of virtual top or bottom quarks which produces a Higgs boson, as high-energy gluons are the most common particles from the proton-proton collisions.
- (ii) Vector boson fusion (VBF): Two quarks or anti-quarks exchange W or Z vector bosons which emits a Higgs boson. The Feynman diagram of this process is shown in figure 2.2 (b). The

initial quarks that radiate the vector bosons are deflected, and they are detected in the form of particle jets as the final state in the detector. The VBF topology is characterized by the presence of the two widely separated jets with high momentum, this distinctive signature gives rise to the great signal acceptance and the large QCD multi-jets background rejection in the corresponding Higgs candidate event selection.

- (iii) Higgs Strahlung (or vector boson associated production, VH): An elementary fermion and the anti-fermion can merge to form a high-energy virtual vector boson (at the LHC, the fermions are quark and anti-quark); if the vector boson has a high energy, it can radiate a Higgs boson by deceleration, as in the bremsstrahlung process. The Feynman diagram of this process is shown in figure 2.2 (c).
- (iv) Top fusion ($t\bar{t}H$): Two gluons each decay into a quark-antiquark pair, a Higgs boson can be produced by the combination of one quark from one pair and one anti-quark from the other pair. Figure 2.2 (d) shows the Feynman diagram of this process.

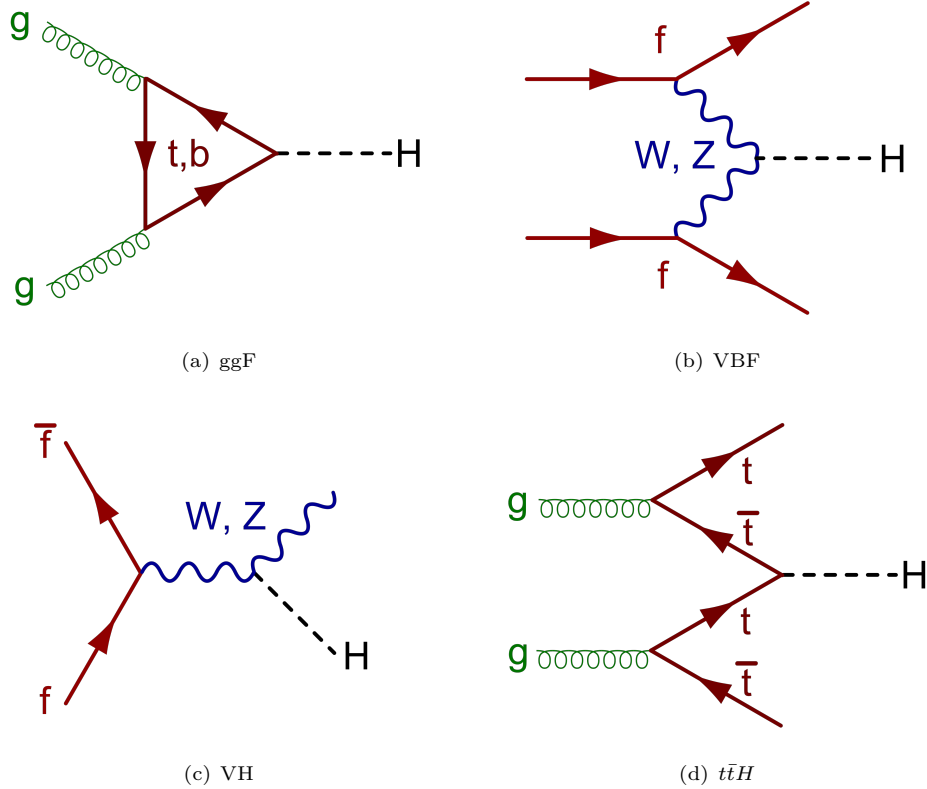


Figure 2.2: The Feynman diagrams for Higgs boson production [21].

The cross sections of Higgs boson production for Higgs mass $m_H = 125 \text{ GeV}/c^2$ and center-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ are summarised in table 2.4. The cross section for the ggF production mode is much larger than for others; VBF production and VH production also have relatively large cross sections, these three production processes are the dominant modes at the LHC. The cross sections as a function of m_H are shown in figure 2.3.

Production Mode	Cross Section [pb^{-1}]	uncertainty
ggF	48.6	+4.6%/-6.7%
VBF	3.78	+0.4%/-0.3%
WH	1.37	+0.5%/-0.7%
ZH	0.88	+3.8%/-3.1%
$t\bar{t}H$	0.50	+5.8%/-9.2%

Table 2.4: The Higgs boson production cross section in the Standard Model [22], computed with $\sqrt{s} = 13$ TeV and $m_H = 125$ GeV/ c^2 . The ggF theory uncertainty and QCD scale uncertainties of other production modes are included.

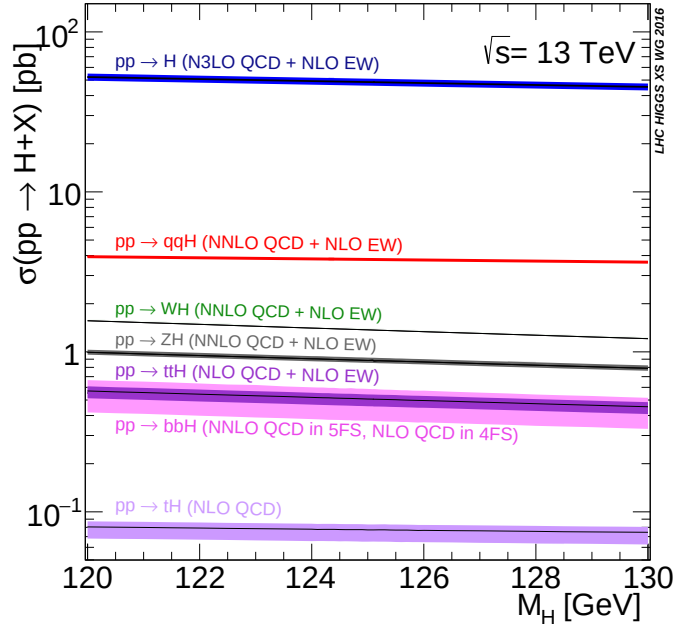


Figure 2.3: The Higgs boson production cross sections as a function of m_H at $\sqrt{s} = 13$ TeV [22].

2.5 Higgs Decays at the LHC

The Higgs boson decays rapidly into other particles. For $m_H = 125$ GeV/ c^2 , its lifetime is extremely short: only about 10^{-22} s [7], according to the SM. Many decay processes of the Higgs boson preserve the SM symmetries. The Higgs boson can interact directly with all the massive elementary particles with a strength proportional to their mass. The fraction of the number of decays in each mode to the total number decays is known as the *branching ratio*. The branching ratio is different in each decay mode and depends on the mass of the particles involved, as well as on other quantities. The decay channels can be classified by the types of output particles: the Higgs boson can split into two fermions or two gauge bosons.

In the decay channels with two fermions, Higgs bosons are more likely to split into heavy fermions due to the energy conservation. For $m_H = 125$ GeV/ c^2 , the most common decay processes involve the heavy particles in the third generation, except top quarks. Although top quarks have the largest mass among elementary particles, the branching ratio of the process $H \rightarrow t\bar{t}$ is zero if the Higgs bosons are not heavy enough. Therefore, the most common decay processes are the $H \rightarrow b\bar{b}$ and $H \rightarrow \tau\bar{\tau}$ according to the SM. The decay processes involving the heavy particles in the second

generation have much smaller branching ratios than those in the third generation. The possible ones are $H \rightarrow c\bar{c}$, $H \rightarrow \mu\bar{\mu}$ and $H \rightarrow s\bar{s}$. Fermions in the first generation are too light, and their branching ratios are negligible among decay channels to two fermions.

The W and Z vector bosons have relatively large masses, and hence the decay channels with two massive vector bosons are also very common. Due to charge conservation, the Higgs boson decays into two massive vector bosons in the way $H \rightarrow W^+W^-$ and $H \rightarrow ZZ$. The $W^\pm$ bosons can subsequently decay into quarks or leptons, and the Z boson also can decay into charged leptons. In other words, the Higgs boson can indirectly decay into four fermions.

Decay channels to two massless gauge bosons are also possible with an intermediate loop of massive vector bosons or virtual heavy quarks. The Higgs boson can decay into two gluons or two photons through a loop of virtual heavy quarks. The decay channel to two photon through a loop of W bosons is also possible but less common. Moreover, one photon could be replaced with the Z boson, resulting in $H \rightarrow Z\gamma$. In the context of detector and reconstruction systems, decay channels with two photons are favoured as they can be detected and measured precisely.

Table 2.5 summarises the branching ratios of the common Higgs boson decay channels, for $m_H = 125 \text{ GeV}/c^2$ and $\sqrt{s} = 13 \text{ TeV}$. Figure 2.4 shows the branching ratios as a function of the Higgs mass m_H . The decay channels with two gauge bosons become the most common ones in the high mass region as heavy Higgs bosons have a larger amount of energy to support the decays.

Decay Channels	Branching ratios	uncertainty
$H \rightarrow f\bar{f}$		
$H \rightarrow b\bar{b}$	5.77×10^{-1}	$+3.21\%/-3.27\%$
$H \rightarrow \tau\bar{\tau}$	6.32×10^{-2}	$+5.71\%/-5.67\%$
$H \rightarrow \mu\bar{\mu}$	2.19×10^{-4}	$+6.01\%/-5.86\%$
$H \rightarrow c\bar{c}$	2.91×10^{-2}	$+12.17\%/-12.21\%$
$H \rightarrow s\bar{s}$	2.46×10^{-4}	$+4.88\%/-4.86\%$
$H \rightarrow VV$		
$H \rightarrow W^+W^-$	2.15×10^{-1}	$+4.26\%/-4.20\%$
$H \rightarrow ZZ$	2.64×10^{-2}	$+4.28\%/-4.21\%$
$H \rightarrow gg$	8.57×10^{-2}	$+10.22\%/-9.98\%$
$H \rightarrow \gamma\gamma$	2.28×10^{-3}	$+4.98\%/-4.89\%$
$H \rightarrow Z\gamma$	1.54×10^{-3}	$+9.01\%/-8.83\%$

Table 2.5: The branching ratios of the Higgs boson decay channels and their uncertainties [22], measured with $m_H = 125 \text{ GeV}/c^2$ at $\sqrt{s} = 13 \text{ TeV}$.

2.6 Lepton Flavour Conservation

In the Standard Model, any decay processes must follow certain conservation laws. According to Noether's theorem, the corresponding symmetries, derived under certain gauge group, are similar to the QED example in equation 2.12. If the symmetry operators commute with the Hamiltonian, the eigenvalues are conserved in the SM processes. However, the conservation of quantum numbers is related to the underlying symmetries of the theory, which means not all those quantum numbers are universally conserved in all interactions. For example, the hypercharge introduced in section 2.2.3 belong to the $SU(3)$ symmetry group and is conserved in strong interactions, but not in weak interactions. The quantum numbers that are strictly conserved in all three SM fundamental

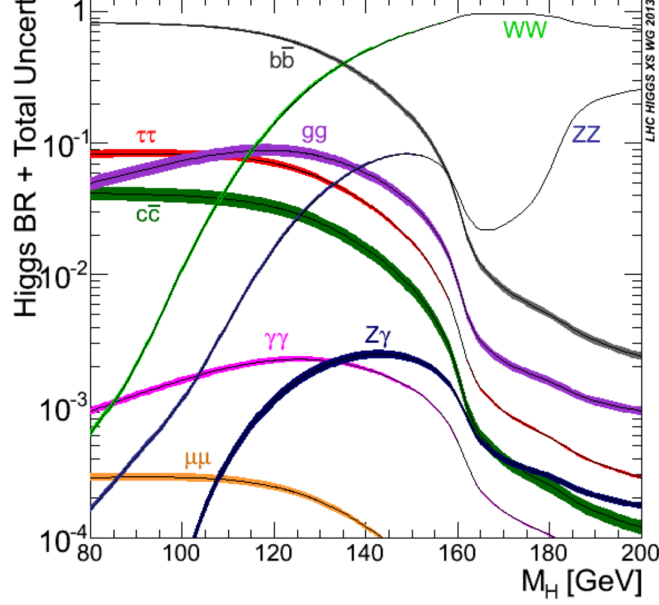


Figure 2.4: The branching ratios of the Higgs boson decay channels as a function of Higgs mass m_H at $\sqrt{s} = 13$ TeV [22].

interactions² are the electric charge Q , *baryon number* B and *lepton number* L . Baryon number and lepton number are defined as:

$$B = \frac{1}{3}(n_q - n_{\bar{q}}), \quad L = n_\ell - n_{\bar{\ell}} \quad (2.48)$$

where q is the number of quarks, $\bar{q}$ is the number of anti-quarks, ℓ is the number of leptons and $\bar{\ell}$ is the number of anti-leptons. The symmetries of renormalizable Lagrangians in the SM precisely conserve the baryon number and lepton number.

In addition to the lepton number, lepton flavour numbers are defined as the number of leptons on each generation or lepton family. For example, the *electron number* L_e represents the number of electrons and electron neutrinos, similarly for *muon number* L_μ and *tau number* L_τ .

The Standard Model predicts lepton flavour conservation, meaning that the number of leptons in each family is conserved. One illustrative example of lepton flavour conservation is the muon decay $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$, where muon number $L_\mu = 1$ and the electron number $L_e = 0$ are conserved before and after the decay.

In the language of gauge theory, the lepton flavour conservation prediction is based on the kinetic term of the Lagrangian for fermions in equation 2.43. Considering ψ as only for leptons this time, the kinetic term can be rewritten in a more general form by assuming that there is no right-handed neutrino, and using the triplet Ψ :

$$\Psi_\alpha = (\nu_{L\alpha}, \psi'_{L\alpha}, \psi'_{R\alpha})^T \quad (2.49)$$

where $\alpha = e, \mu, \tau$ and ψ' stands for charged leptons. To avoid breaking the gauge symmetry, the Higgs field also should be included in the Lagrangian as in equation 2.42. The kinetic term of the Lagrangian is invariant under the phase shift of each flavour $\mathcal{L}'_\ell \rightarrow e^{-i\theta} \mathcal{L}_\ell$, $\mathcal{L}'_\ell = \mathcal{L}_\ell$ and $\theta \rightarrow \theta + \delta\theta(x)$.

²Weak isospin is conserved in the three fundamental interactions, but it is not conserved in the interactions with the Higgs field.

According to Noether's theorem, there is a conserved current for the lepton flavour conservation:

$$j_\alpha^\mu = \bar{\nu}_{L\alpha} \gamma^\mu \nu_{L\alpha} + \bar{\psi}'_{L\alpha} \gamma^\mu \psi'_{L\alpha} + \bar{\psi}'_{R\alpha} \gamma^\mu \psi'_{R\alpha} \quad (2.50)$$

To explore lepton flavour conservation in Higgs boson decays, which is more relevant to the objective of this thesis, one needs to consider the Yukawa Lagrangian in equation 2.47: inserting the SU(2) left-handed doublet and right-handed singlet for leptons, and expanding equation 2.47 in terms of the state $\psi(x)$ in equation 2.46:

$$\mathcal{L}_{Yukawa} = -Y_f(v+h)(\bar{f}_L f_R + \bar{f}_R f_L) + h.c. \quad (2.51)$$

where Y_f is the Yukawa constant for fermions, and f stands for e , μ and τ in this case. The term with $Y_f v$ gives masses to the fermions, and the neutrino term vanishes, reflecting the fact that neutrinos are massless in the Standard Model. Rewriting the Lagrangian in terms of the mass basis gives:

$$\mathcal{L}_{Yukawa} = -m_i \bar{f}_L^i f_R^i - Y_{ij}(\bar{f}_L^i f_R^j)h + h.c. + \dots \quad (2.52)$$

where $m_i = Y v$, and $Y_{ij} = (m_i/v)\delta_{ij}$ in the SM, v is taken as the electroweak scale of 246 GeV. The lepton flavour symmetry is preserved in the SM without breaking the gauge symmetries.

However, the observation of neutrino oscillations indicates that lepton flavour is not conserved in the neutral sector, and neutrinos have mass. The details are discussed in section 3.1.1.

Chapter 3

Beyond the Standard Model and Lepton Flavour Violation

The Standard Model, together with the general relativity, provides a remarkable framework for explaining most of the physical processes observed in nature. Despite this tremendous success, the SM itself is inconsistent with general relativity, and many observations remain unexplained by the SM. These include the mysteries of dark energy and dark matter, the matter-antimatter asymmetry, and others. Some current challenges for the SM are discussed in section 3.1, and the observation of neutrino oscillations is discussed in more detail in section 3.1.1.

The discovery of the Higgs boson in 2012 at CERN [1, 2] marked the completion of the SM of particle physics. The SM is so consistent with observations that almost all the possible deviations due to new physics have to be suppressed by making the energy scale of new physics larger than about 10 TeV. However, if one goes back in the history of particle physics to the 1970s, particle interactions can be described by the model involving QED, QCD and Fermi theory at that time. This model, however, is an *effective field theory* that is valid only up to an energy scale of order 100 GeV (W and Z boson mass) [7]. Beyond this energy scale, more precise measurements and higher energy processes revealed limitations in the theory. The primary issue is related to the renormalizability of the theory. The SM is thus currently seen as an effective field theory, motivated by the unexplained theoretical and experimental problems.

An effective field theory is a general quantum field theory that does not have the restriction of renormalizability, it is valid only up to certain energy scale Λ . In contrast, *fundamental theories* can be used to describe nature at any energy scale. Certainly, the SM is not a candidate for a fundamental theory. Despite being renormalizable, the SM describes particle physics in the low-energy limits of fundamental theories. A possible candidate for the fundamental theory could be *string theory*, which does not require a cutoff for its validity.

Renormalization is a procedure used to address divergences. A theory is non-renormalizable if the Lagrangian contains field operators with dimension higher than four $(E/\Lambda)^{d-4}$, and an infinite number of counterterms are required to cancel the divergences. This results in an infinite number of free parameters and leads to a loss of predictive power. While the SM is renormalizable, the theory of general relativity is not, as the coupling contains a negative dimension, $G = \hbar c/M_P^2$, where M_P is the Planck mass with a scale of 10^{19} GeV. The inconsistency between quantum gravity and the SM indicates the SM is not a complete theory and suggests the existence of new physics beyond

it. However, renormalizability should not be a constraint on a theory. Rather, the coefficients in the high-dimensional field operators should be suppressed by the inverse power of Λ , such that $(E/\Lambda)^{d-4} \ll 1$. Interactions could be physically described by these additional terms.

In this chapter, experimental evidence for new physics beyond the SM below M_P is introduced in section 3.1 and section 3.1.1, and the theoretical reasons to believe that there should be new physics at the collider-accessible TeV scale are introduced in section 3.3.

3.1 Challenges for the Standard Model

An effective field theory is considered natural if all its parameters are of order unity in the units of its cutoff Λ . The natural parameters should be of the same order as the value estimated from the dimensional analysis, which take the form Λ^{4-d} , where d is the dimension of the field operator. However, in the Standard Model, some parameters are vastly different from the required scales, contradicting the assumption of *naturalness*. These parameters suggest that there should be new symmetries or new physics beyond the SM and give rise to the most famous puzzles which remain unsolved in the SM:

- **The Hierarchy Problem**

New physics may be found in the Higgs sector, as the Higgs mass is quadratically sensitive to the energy scale of new physics. In the SM, the mass corrections to a scalar particle are quadratically divergent, as there is no known symmetry to suppress the divergence, unlike the chiral fermions and gauge particles. For example, in QED, the Lagrangian preserves a chiral symmetry as the fermion masses go to zero:

$$\psi \rightarrow e^{i\alpha\gamma_5}\psi, \quad \psi \rightarrow e^{i\alpha\tau/2\gamma_5}\psi \quad (3.1)$$

The first transformation is for U(1) and the second is for SU(2). This symmetry guarantees that all the radiative corrections will be zero as $m \rightarrow 0$.

The Higgs potential in the Standard Model is given in the second term of equation 2.45, and the Higgs mass with radiative correction δm_H is given by $m_H^2 = \lambda\eta^2/2 + \delta m_H^2$. There are three main types of radiative corrections to the Higgs mass: the Yukawa coupling with top quark loop, the gauge boson loop, and the Higgs self-coupling. These corrections are proportional to the energy cutoff Λ as [23]:

$$\delta m_H^2 \simeq \frac{3}{4\pi^2} \left(-\frac{1}{2}y_t^2 + \frac{g^2}{12} + \frac{\lambda^2}{12} \right) \Lambda^2 \quad (3.2)$$

where y_t is the top-quark Yukawa coupling and g is the SU(2) coupling. The radiative corrections should have the same order as the typical energy scale of the fields; or in other words, by naturalness, $\delta m_H^2 < m_H^2$. On the other hand, as the SM is an effective field theory, the energy should be cut off below the scale where new physics arises. If one tries to insert the Planck scale of 10^{19} GeV, the corrections to the Higgs mass will be gigantic compared to the Higgs mass itself. The observed Higgs mass $m_H = 125$ GeV and equation 3.2 imply that the SM, as an effective field theory regarding naturalness, is valid only up to a few TeV, which is a collider-accessible scale.

This hierarchy problem can be resolved by introducing new symmetries to suppress the divergence; this approach, known as the *supersymmetry*, is discussed in section 3.3.2. Another solution to this puzzle is the *composite Higgs model*, which assumes that the Higgs boson is a composite particle; the details are discussed in section 3.3.3. Other potential solutions are models with extra dimension; these models also predict the LFV decays of the Higgs boson.

Other puzzles concerning the naturalness problem are the *strong CP problem* and the *cosmological constant problem*. The former puzzle even occurs at the dimension-four operators, and hence it is related to a dimensionless parameter θ . In the SM, CP-violation could occur in the strong interactions, but no experimental observation have been occurred so far. In the QCD Lagrangian, CP-violation is represented by a term with the parameter θ that can take any value between zero and π and is proportional to the neutron electric dipole moment (nEDM); however, experimental measurements on the nEDM require $\theta < 10^{-10}$ [24]. The latter puzzle refers to the large discrepancy between the observed vacuum energy density and the theoretical prediction.

In particle physics, scientists' confidence in new discoveries is expressed in terms of its statistical significance, which in term, is expressed in terms of a number of standard deviations (or sigma). At least five standard deviations are required to call a finding as *discovery*. In modern physics, the precision of the measurements on some physical parameters are largely enhanced thanks to the upgrades of experimental apparatus. After the precision was increased, some measurements were found to contradict the SM with a high statistical significance:

- **Muon g-2 Experiment**

In April 2021, Fermilab published their results of the *Muon g-2* experiment, obtained using the data collected from 2018, which provided precise measurements of the anomalous magnetic moment of muon. The measured value $a_\mu = 0.00116592061(41)$ [25], exceeded the SM prediction by 4.2σ . At the time of writing (August 2023), Fermilab just announced their latest results, obtained using the data collected from 2018 to 2020 [26], which is $a_\mu = 0.00116592055(24)$ with a precision of 0.20 ppm.

For an elementary particle with intrinsic angular momentum $\mathbf{S}$ and electric charge q , its magnetic moment $\boldsymbol{\mu}$ can be expressed as $\boldsymbol{\mu} = g \frac{q}{2m} \mathbf{S}$, where g is the g-factor. According to the Dirac equation, the g-factor for spin- $\frac{1}{2}$ particles is predicted to be $g = 2$. The anomalous magnetic moment a represents the contribution to the magnetic moment of the particle, and it can be expressed as the fractional deviation from the Dirac's prediction, $a = \frac{g - 2}{2}$.

The anomalous magnetic moment is significant as the precise measurements provide a stringent test of the SM. Fermilab's results are just slightly below the 5σ threshold, providing compelling evidence of new physics.

Scientists have been trying to explain the discrepancy in the anomalous magnetic moment of the muon using theoretical models such as supersymmetry [27], discussed in section 3.3.2, and the two Higgs doublet model [28–30], discussed in section 3.3.1. These models also predict LFV decays of the Higgs boson, which provides a strong motivation for the Higgs LFV searches; the LFV results can put strong constraints on these models if the muon g-2 anomaly is caused by new physics.

Lepton flavour violating couplings of the Higgs boson may have contributions to the muon g-2 anomaly. Figure 3.1 shows the Feynman diagram of the process involving $\mu - \tau$ flavour-

violating couplings. The specific flavour-violating contribution to the muon g-2 anomaly could be attributed to the τ -Higgs loop [31]:

$$\Delta a_\mu \equiv a_\mu^{exp} - a_\mu^{SM} \simeq \frac{Re(Y_{\mu\tau}Y_{\tau\mu})}{8\pi^2} \frac{m_\mu m_\tau}{2m_h^2} (2\ln \frac{m_h^2}{m_\tau^2} - 3) \quad (3.3)$$

where $Y_{\mu\tau}$ and $Y_{\tau\mu}$ are the Yukawa coupling terms. Substituting the Δa_μ measured at Fermilab and the values of the masses into equation 3.3:

$$Re(Y_{\mu\tau}Y_{\tau\mu}) \simeq 2.4 \times 10^{-3} \quad (3.4)$$

which is consistent with the bounds and constraints reported in [32]. Hence, the Fermilab measurement could be explained by the flavour-violating Yukawa couplings with the size given in equation 3.4. Comparing to the value of $Re(Y_{\mu\tau}Y_{\tau\mu})$ obtained by using the Δa_μ measurement from the E821 experiment at Brookhaven National Laboratory (BNL) [33] ($Re(Y_{\mu\tau}Y_{\tau\mu}) = 2.7 \times 10^{-3}$), one can notice that the Fermilab result, with higher precision, is closer to the bound reported in [32], as expected.

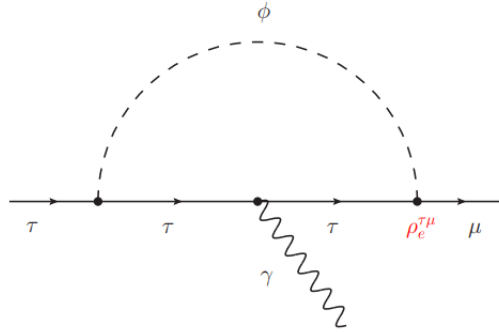


Figure 3.1: The contribution to muon g-2 anomaly from the Higgs mediated $\tau \rightarrow \mu\gamma$ with one loop [34], ϕ can be the neutral Higgs boson h and H, or the CP-odd pseudoscalar A.

Another recent publication from the Collider Detector at Fermilab (CDF) Collaboration reported a measurement of the W boson mass with high precision. In April 2022, the CDF collaboration announced that the mass of the W boson is 80433 ± 9 MeV [35] which differs from the SM predictions by 7σ . This measurement largely catches particle physicists' attention, as it would constitute evidence of physics beyond the SM if confirmed. However, ATLAS has published its latest measurements on the W boson mass with the highest precision, the published value is $m_W = 80360 \pm 16$ MeV [36], which is compatible with the SM and the uncertainty is improved by 15% comparing to the previous ATLAS results.

3.1.1 Neutrino Oscillations

One of the strongest motivations for LFV searches is the experimental observation of neutrino oscillations [3, 4]. These observations indicate that neutrinos have mass and LFV exists in nature, which is in contradiction with the SM predictions.

Since neutrinos have zero electric charge and interact weakly, neutrinos from the Sun are the ideal candidates for studying the interior of the Sun. In 1960s, Ray Davis' Homestake experiment

reported the first measurements on the solar neutrinos [37]. The total accumulation of the solar neutrinos was only about one-third of the Standard Solar Model prediction, this deficit leads to the long-existing solar neutrino problem.

Bruno Pontecorvo presented an explanation for the solar neutrino problem in 1968 [38], he proposed that the flavour of a neutrino can vary between three known states as it travels in space. This phenomenon is known as *neutrino oscillations*, and can be described in terms of mixing states in quantum mechanics. Neutrino oscillations were observed in Super-Kamiokande and the Sudbury Neutrino observatories, and their confirmation of neutrino oscillations led to awarding of the 2015 Nobel Prize for Physics.

Neutrinos can spontaneously change flavour spontaneously during propagation. The mass-eigenstates can be expressed in the form of orthogonal linear combination of different flavour states:

$$|\nu_1\rangle = \cos\theta |\nu_\beta\rangle - \sin\theta |\nu_\alpha\rangle, \quad |\nu_2\rangle = \sin\theta |\nu_\beta\rangle + \cos\theta |\nu_\alpha\rangle \quad (3.5)$$

where θ is the mixing angle. The probability of a neutrino with flavour β to convert into flavour α after time t can be shown to be:

$$P_{\nu_\alpha \rightarrow \nu_\beta} = |\langle \nu_\beta | \nu_\alpha \rangle|^2 = \sin^2(2\theta) \sin^2\left(\frac{E_2 - E_1}{2\hbar}t\right) \quad (3.6)$$

As neutrinos propagate with untrarelativistic speed, the energy difference between the two mass-eigenstates can be approximated as:

$$E \simeq |\mathbf{p}c| + \frac{m^2 c^3}{2|\mathbf{p}|}, \quad E_2 - E_1 \simeq \frac{m_2^2 - m_1^2}{2E} c^4 \quad (3.7)$$

Hence, the probability of the flavour conversion in equation 3.6 can be expressed in terms of the mass difference. The experimental measurements of this probability function have confirmed the evidence of neutrino oscillations and hence that neutrinos have mass. However, current experiments cannot provide meaningful measurements of the mass of any individual neutrino flavour due to low sensitivities.

Neutrino oscillations reveal that neutrinos have mass, and flavour transitions can be represented in terms of the mass-eigenstates and a mixing matrix. The three-flavour oscillations are described by the mixing matrix for the leptons, analogous to the CKM matrix for quarks. This matrix is also known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (3.8)$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. Each component in the mixing matrix represents an amplitude of mass-eigenstate in the construction of a flavour-eigenstate.

3.2 Lepton Flavour Violation

In the Standard Model, lepton flavour violation is forbidden as neutrinos are predicted to be left-handed and massless, resulting in no mixing matrix for leptons. However, the observation of

neutrino oscillations proves the existence of LFV in the neutral sector and opens up the possibility for the charged lepton flavour violation.

Using the PMNS matrix in equation 3.8 and the measured mixing angles, the branching ratio of the most searched LFV decays, induced by neutrino oscillations, can be calculated. For example, the branching ratio for $\mu \rightarrow e\gamma$ at one-loop is given by:

$$\mathcal{B}(\mu \rightarrow e\gamma) = \frac{3\alpha}{32\pi} \left| \sum_{i=2,3} U_{ei}^* U_{\mu i} \frac{\Delta m_{i1}^2}{m_W^2} \right|^2 \sim 10^{-54} \quad (3.9)$$

where U_{ei} and $U_{\mu i}$ are the elements of the PMNS matrix, and m_W is the mass of the W boson. The branching ratio is strongly suppressed by the term $\frac{\Delta m_{i1}^4}{m_W^4}$, resulting in an extremely small value that is beyond experimental reach. The upper limit on $\mathcal{B}(\mu \rightarrow e\gamma) = 4.2 \times 10^{-13}$ at 90% confidence level was measured at the *Mu to E Gamma* experiment [39]. The new physics contributions to this recent measured $\mu \rightarrow e\gamma$ upper bound and the muon g-2 anomaly were studied in [40] by using the GIM mechanism and a leptoquark model.

There are other promising LFV searches, such as the $\mu^+ \rightarrow e^+e^-e^+$ decay with an upper limit on $\mathcal{B}(\mu^+ \rightarrow e^+e^-e^+) < 1.0 \times 10^{-12}$ at 90% C.L. [41], measured in the SINDRUM experiment in 1988. The 2022 Particle Data Group summarised the branching ratio of LFV decays of K mesons [7]. The measurements of the branching ratio of the LFV tau decays: $\mathcal{B}(\tau \rightarrow 3\ell_i)$ and $\mathcal{B}(\tau \rightarrow \ell_i \ell_j \ell_j)$ ($i, j = e, \mu$) were done by the Belle collaboration [42]. The $\mathcal{B}(\tau \rightarrow \mu\gamma) < 4.4 \times 10^{-8}$ and $\mathcal{B}(\tau \rightarrow e\gamma) < 3.3 \times 10^{-8}$ at 90% C.L. provide indirect constraints on the upper bounds of Higgs LFV decays; the measurements were done by the BaBar collaboration [43]. Prior searches for LFV decays of the Higgs boson are summarised in section 3.2.2.

3.2.1 LFV Decays of the Higgs Boson

Searches for new physics in the Higgs sector have garnered interest in recent high energy physics experiments. Also, precise measurements of its properties can provide much deeper insights into particle physics. Among the searches, the direct searches for LFV decays of the Higgs boson is particularly attractive. This motivation stems from both theoretical and observational reasons, as well as from the experimental measurements conducted by CMS and ATLAS, as discussed in section 3.2.2.

The Higgs couplings in the SM are diagonal according to equation 2.52. However, in extensions of the SM (BSM), the Yukawa term can deviate from the form $Y_{ij} = (m_i/v)\delta_{ij}$ due to the presence of LFV decays. LFV decays of the Higgs boson can occur at tree level through the Yukawa interactions [31], involving decay channels such as $H \rightarrow e\mu$, $H \rightarrow e\tau$, and $H \rightarrow \mu\tau$:

$$\mathcal{L}_{HLFV} \supset -Y_{\mu\tau} h \bar{\mu}_L \tau_R - Y_{\tau\mu} h \bar{\tau}_L \mu_R - Y_{e\tau} h \bar{e}_L \tau_R - Y_{\tau e} h \bar{\tau}_L e_R - Y_{\mu e} h \bar{\mu}_L e_R - Y_{e\mu} h \bar{e}_L \mu_R + h.c. \quad (3.10)$$

The non-diagonal Yukawa coupling matrix elements can be derived from the branching ratios of the Higgs LFV decays:

$$|Y_{\ell\tau}|^2 + |Y_{\tau\ell}|^2 = \frac{8\pi\Gamma_H(SM)}{m_H} \frac{\mathcal{B}(H \rightarrow \ell\tau)}{1 - \mathcal{B}(H \rightarrow \ell\tau)} \quad (3.11)$$

where $\Gamma_H(SM) = 4.07 \text{ MeV}$ [44] is the SM prediction for the total decay width of the Higgs boson.

The most stringent indirect constraints on the Higgs LFV couplings are derived from the upper limits on the processes $\mu \rightarrow e\gamma$, $\tau \rightarrow \mu\gamma$ and $\tau \rightarrow e\gamma$. By using the result $\mathcal{B}(\mu \rightarrow e\gamma)$, a bound on the $\mathcal{B}(H \rightarrow e\mu) < \mathcal{O}(10^{-8})$ can be derived [31, 32].

3.2.2 Prior Higgs LFV Searches

The most stringent bounds on lepton flavor-violating (LFV) decays of the Higgs boson arise from direct searches conducted by the ATLAS and CMS experiments using proton-proton collision data from the LHC. In the Run 1 operation at the LHC, both CMS and ATLAS collected data for two years, from 2011 to 2012; the dataset from the Run 2 operation used data collected between 2015 and 2018. No significant excess has been observed in the recent experimental results.

The first dedicated search was reported by the CMS collaboration in 2015, using the data collected in 2012 with a centre of mass energy of $\sqrt{s} = 8$ TeV and corresponding to an integrated luminosity of 19.7 fb^{-1} . Only searches on the $H \rightarrow \mu\tau$ channel were performed, and a slight excess of signal events was observed with a 2.4σ significance. An upper limit on the $\mathcal{B}(H \rightarrow \mu\tau) < 1.51\%$ was set at 95% C.L., constraining the Yukawa coupling $\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 3.6 \times 10^{-3}$ [45].

A similar search was performed by ATLAS, utilizing data collected by the ATLAS detector corresponding to an integrated luminosity of 20.3 fb^{-1} and a centre-of-mass energy $\sqrt{s} = 8$ TeV. The analysis was only performed on the $H \rightarrow \mu\tau_{had}$ decay channel, where τ_{had} is the hadronic decay of the *tau*-lepton. No excess of signal events was found, and an upper limit on the $\mathcal{B}(H \rightarrow \mu\tau) < 1.85\%$ at 95% C.L. was set [46], which is consistent with the CMS result. A second analysis was performed by ATLAS using the dataset with the same luminosity and centre of mass energy as the previous ATLAS result, the results were published in 2017. The searches were performed through the decay channels $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$, with the τ -lepton decaying either to hadrons or to leptons (electron and muon). Upper limits on $\mathcal{B}(H \rightarrow e\tau) < 1.04\%$ and $\mathcal{B}(H \rightarrow \mu\tau) < 1.43\%$ at 95% C.L. were set [47].

CMS also performed searches using the partial Run 2 dataset collected in 2016, with a centre-of-mass energy $\sqrt{s} = 13$ TeV and an integrated luminosity of 35.9 fb^{-1} . The observed upper limits on $\mathcal{B}(H \rightarrow e\tau) < 0.61\%$ and $\mathcal{B}(H \rightarrow \mu\tau) < 0.25\%$ at 95% C.L. were set [48], and these limits were used to constrain the Yukawa couplings of $\sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 2.26 \times 10^{-3}$ and $\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 1.43 \times 10^{-3}$. In comparison, the partial Run 2 analysis was also performed by ATLAS using the dataset collected from 2015-2016 at $\sqrt{s} = 13$ TeV and with an integrated luminosity of 36.1 fb^{-1} . The ATLAS collaboration reported 95% C.L. upper limits $\mathcal{B}(H \rightarrow e\tau) < 0.47\%$ and $\mathcal{B}(H \rightarrow \mu\tau) < 0.28\%$ [49], which are in consistent with the CMS results.

The latest results reported by the CMS collaboration were reported recently, using the dataset collected from 2016-2018 at $\sqrt{s} = 13$ TeV and with an integrated luminosity of 137 fb^{-1} . The observed upper limits on the branching fractions are $\mathcal{B}(H \rightarrow e\tau) < 0.22\%$ and $\mathcal{B}(H \rightarrow \mu\tau) < 0.15\%$, and the limits on the Yukawa couplings are $\sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 1.35 \times 10^{-3}$ and $\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 1.11 \times 10^{-3}$ [50].

The most stringent ATLAS results to date were reported recently using the full Run 2 dataset with $\sqrt{s} = 13$ TeV and an integrated luminosity of 138 fb^{-1} . Two background estimation methods are employed in the full Run 2 analysis: the MC-template method and the symmetry method; the latter is more sensitive to the branching ratio difference. The details of the results are discussed in section 14, in short, the observed upper limits on the branching fractions are $\mathcal{B}(H \rightarrow e\tau) < 0.20\%$ and $\mathcal{B}(H \rightarrow \mu\tau) < 0.18\%$, and the best-fit branching ratio difference, $\mathcal{B}(H \rightarrow e\tau) - \mathcal{B}(H \rightarrow \mu\tau)$,

measured in the channels where the τ –lepton decays to leptons is compatible with zero within 2.5σ [51].

All the results of these historical searches are summarised in table 3.1. The CMS and ATLAS results with similar $\sqrt{s}$ and luminosity are consistence as expected, lower upper limits on the branching fractions and stronger constrain on the Yukawa couplings are obtained with higher energy and luminosity.

Collaboration (publication year)	$\sqrt{s}$ [TeV]	Luminosity [fb ⁻¹]	Upper limit on $\mathcal{B}(H \rightarrow e\tau)$	Upper limit on $\mathcal{B}(H \rightarrow \mu\tau)$	$\sqrt{ Y_{e\tau} ^2 + Y_{\tau e} ^2}$	$\sqrt{ Y_{\mu\tau} ^2 + Y_{\tau\mu} ^2}$
CMS (2015) [45]	8	19.7	-	< 1.57%	-	< 3.60×10^{-3}
ATLAS (2015) [46]	8	20.3	-	< 1.85%	-	-
ATLAS (2017) [47]	8	20.3	< 1.04%	< 1.43%	-	-
CMS (2018) [48]	13	35.9	< 0.61%	< 0.25%	< 2.26×10^{-3}	< 1.43×10^{-3}
ATLAS (2020) [49]	13	36.1	< 0.47%	< 0.28%	-	-
CMS (2021) [50]	13	137	< 0.22%	< 0.15%	< 1.35×10^{-3}	< 1.11×10^{-3}
ATLAS (2023) [51]	13	138	< 0.20%	< 0.18%	< 1.40×10^{-3}	< 1.20×10^{-3}

Table 3.1: Summary of previous results from the CMS and ATLAS collaborations.

3.3 Higgs LFV predictions in BSM Theories

There is no LFV effect at tree-level in the SM; the new physics contributions can come from the operators with dimension higher than four. If Λ_{BSM} is the scale of physics beyond the Standard Model, the operators with higher dimension will be suppressed by progressively larger powers of Λ_{BSM} . The general form of the effective Lagrangian is:

$$\mathcal{L}_{eff} = \mathcal{L}_{SM} + \sum_i \frac{C_i^{(5)}}{\Lambda} O_i^5 + \sum_i \frac{C_i^{(6)}}{\Lambda^2} O_i^6 + \dots \quad (3.12)$$

where O_i^d is the dimension-d operator and $C_i^{(d)}$ is the associated Wilson coefficient. The LFV effects can be described by $\mathcal{L}_{eff}$ with a scale of new physics Λ in a model-independent way, which disappear as $\Lambda \rightarrow \infty$.

In the SM, there is only one dimension-five operator known as the Weinberg operator. This operator generates left-handed neutrino masses of the Majorana type after electroweak symmetry breaking and induces LFV vertices. The dimension-five operator is a dipole operator at tree-level [52]:

$$O_{ij}^5 = \bar{\tilde{L}}_i \tau_a L_j \tilde{\Phi} \tau_a \Phi \quad (3.13)$$

where i, j ($=1, 2, 3$) denote the flavor indices, L is the lepton doublet and Φ is the Higgs doublet. However, it has been shown that this operator does not induce the LFV decays of the neutral Higgs boson [53].

Dimension-six operators can also induce the Higgs LFV interactions; some of them can arise at tree-level or as one-loop corrections to the fundamental theory. For operators that only generate the LFV interactions of Z boson and Higgs bosons at tree-level [53]:

$$\begin{aligned}
O_{ij}^6 &= i(\Phi^\dagger D^\mu \Phi)(\bar{\psi}_{Ri} \gamma_\mu \psi_{Rj}), \\
O_{ij}^{6(1)} &= i(\Phi^\dagger D^\mu \Phi)(\bar{L}_i \gamma_\mu L_j), \\
O_{ij}^{6(3)} &= i(\Phi^\dagger \tau^a D^\mu \Phi)(\bar{L}_i \gamma_\mu \tau^a L_j)
\end{aligned} \tag{3.14}$$

And for operators that arise as one-loop effects [53]:

$$\begin{aligned}
O_{ij}^6 &= (\bar{L}_i D^\mu \psi_{Rj}) D_\mu \Phi, \\
O_{ij}^6 &= (\bar{D}^\mu L_i \psi_{Rj}) D_\mu \Phi
\end{aligned} \tag{3.15}$$

These operators can be used to derive indirect constraints on the bounds of $H \rightarrow \ell_i \ell_j$ from the bounds of $Z \rightarrow \ell_i \ell_j$. However, any gauge-invariant operators that induce the Higgs LFV vertices are reducible to the Yukawa-type operator [34] in equation 3.16.

The only dimension-six operator that induces LFV decays of the Higgs boson is the Yukawa-type operator [53]:

$$O_{Yukawa}^6 = (\Phi^\dagger \Phi)(\bar{L}_i \psi_{Rj} \Phi) \tag{3.16}$$

With the SM Yukawa Lagrangian in equation 2.47 and the dimension-six effective Lagrangian with the Yukawa-type operator, one can write down the effective Lagrangian with Higgs LFV effects:

$$\mathcal{L}_{HLFV} = -\bar{L} \Gamma \psi_R \Phi + \frac{C^{(6)}}{\Lambda^2} (\Phi^\dagger \Phi)(\bar{L} \psi_R \Phi) + h.c. \tag{3.17}$$

where $\mathcal{L}_{SM}^Y = -\bar{L} \Gamma \psi_R \Phi$ and Γ is a complex diagonal matrix. After electroweak symmetry breaking, the lepton doublet L reduces to the charged lepton singlet ψ and the Higgs doublet Φ reduces to the form in equation 2.46, hence:

$$\begin{aligned}
\mathcal{L}_{HLFV} &= -\frac{1}{\sqrt{2}} \bar{\psi}_L \Gamma \psi_R (v + h) + \frac{C^{(6)}}{2\sqrt{2}\Lambda^2} (v + h^\dagger)(v + h)(\bar{\psi}_L \psi_R)(v + h) + h.c. \\
&= \bar{\psi}_L \left[\frac{v}{\sqrt{2}} (\Gamma - C^{(6)} \frac{v^2}{2\Lambda^2}) + \frac{h}{\sqrt{2}} (\Gamma - C^{(6)} \frac{v^2}{2\Lambda^2} - C^{(6)} \frac{v^2}{2\Lambda^2} - C^{(6)} \frac{v^2}{2\Lambda^2}) \right] \psi_R + h.c. \\
&= \bar{\psi}_L \left[\frac{v}{\sqrt{2}} (\Gamma - C^{(6)} \frac{v^2}{2\Lambda^2}) + \frac{h}{\sqrt{2}} (\Gamma - 3C^{(6)} \frac{v^2}{2\Lambda^2}) \right] \psi_R + h.c.
\end{aligned} \tag{3.18}$$

By comparing this Lagrangian to the Yukawa Lagrangian in equation 2.52, one can write down the charged lepton mass matrix $\mathcal{M}$ and the Higgs boson couplings $\mathcal{Y}$ to a pair of leptons, similar to the Y_{ij} in equation 2.52:

$$\mathcal{M} = \frac{v}{\sqrt{2}} (\Gamma - C^{(6)} \frac{v^2}{2\Lambda^2}), \quad \mathcal{Y} = \frac{1}{\sqrt{2}} (\Gamma - 3C^{(6)} \frac{v^2}{2\Lambda^2}) \tag{3.19}$$

In mass basis with the unitary matrices V_L and V_R that diagonalized $\mathcal{M}$:

$$V_L^\dagger \mathcal{M} V_R = \hat{\mathcal{M}} = \text{diag}(m_e, m_\mu, m_\tau) \tag{3.20}$$

The Yukawa coupling term in equation 2.52 can be rewritten effectively as:

$$\begin{aligned}
Y_{ij} &= V_L^\dagger \mathcal{Y} V_R = \frac{\hat{\mathcal{M}}}{v} - \frac{v^2}{\sqrt{2}\Lambda^2} V_L^\dagger C^{(6)} V_R \\
&= \frac{m_i}{v} \delta_{ij} + \frac{v^2}{\sqrt{2}\Lambda^2} \hat{\lambda}'
\end{aligned} \tag{3.21}$$

where $\hat{\lambda}' = -V_L^\dagger C^{(6)} V_R$, which is a non-diagonal matrix. It has been proved that $Y_{ij} \neq \frac{m_i}{v} \delta_{ij}$ in the BSM, and there are LFV contributions from the effective Lagrangian with dimension-six Yukawa-type operator. The effective Yukawa coupling involves a quantum fluctuation $\frac{v^2}{\sqrt{2}\Lambda^2} \hat{\lambda}'$, if one wants to avoid the tuning or hierarchy problem, similar to the Higgs boson mass, a condition on the couplings is needed [31]

$$|Y_{ij} Y_{ji}| \lesssim \frac{m_i m_j}{v^2} \tag{3.22}$$

The branching ratios of the LFV decays of the Higgs boson can be derived in terms of the effective Yukawa couplings [34]:

$$\mathcal{B}(h \rightarrow \ell_i \ell_j) = \frac{m_h}{8\pi\Gamma_H(SM)} (|Y_{ij}|^2 + |Y_{ji}|^2) \tag{3.23}$$

The naturalness limit in equation 3.22 puts a constraint on the branching fractions of the Higgs LFV decays, while the real value contributes to the anomalous magnetic moment of the muon, as in equation 3.3. One can easily verify that the results in table 3.1 match the bounds. A stronger constraint on the bounds of $H \rightarrow e\tau$ from the naturalness limit is expected.

Besides the indirect constraints on the Higgs LFV couplings from the bounds of $Z \rightarrow \ell_i \ell_j$, common indirect constraints could also come from the $\ell_i \rightarrow \ell_j \gamma$ decays as discussed in section 3.2. Considering the dimension-five operator in equation 3.13, the photonic dipole operator at loop-level is [34]:

$$O_{e\gamma} = \bar{\psi}_L \sigma^{\mu\nu} \psi_R F_{\mu\nu} \tag{3.24}$$

where $\sigma^{\mu\nu}$ are the Pauli matrices. This operator has contributions from the $\ell_i \rightarrow \ell_j \gamma$ LFV decays. The branching ratios can be derived in terms of the Wilson coefficients $g_{e\gamma}$ [34]:

$$\mathcal{B}(\ell_i \rightarrow \ell_j \gamma) = \frac{m_i^3}{4\pi v^2 \Gamma_i} (|g_{ij}|^2 + |g_{ji}|^2) \tag{3.25}$$

where Γ_i is the total decay width of the charged lepton ℓ_i . As discussed in section 3.2, indirect constraints on the objective Higgs LFV decays can come from the LFV decays induced by the operator in equation 3.24. If the couplings g_{ij} are close to or greater than the effective Yukawa couplings Y_{ij} , the Higgs LFV decays will be strongly constrained.

The indirect or direct constraints on the Higgs LFV decays obtained from the model-independent studies are matched with the experimental results shown in table 3.1 in general. This opens the possibilities for new physics effects within the reach of the ATLAS and CMS detectors. The Standard Model needs to be modified in order to explain the hierarchy problem, the observation of neutrino oscillations, the experimental results of Muon g-2 and the anomalous W boson mass, and so on. In fact, many physicists have been working on new models that can accommodate some of these BSM findings even before the pre-LHC era. In this section, several BSM theories that can address the SM's challenging problems and predict the Higgs LFV decays at the same time will be introduced.

However, one should notice that the searches in this thesis do not depend on any theoretical model as the benchmark; instead, these models strongly motivate the Higgs LFV searches.

3.3.1 Two-Higgs-Doublet Model

The discovered Higgs boson, consistent with a doublet structure, could be explained by the Two-Higgs-Doublet Model (2HDM), which extends the SM by introducing an additional Higgs doublet. In this model, instead of a single Higgs doublet ψ as in equation 2.45, the Higgs mechanism involves two Higgs doublets, φ_1 and φ_2 . Many BSM theories, such as the Minimal Supersymmetric Standard Model (MSSM), discussed in section 3.3.2, also require additional Higgs doublets to address the hierarchy problem.

Different types of 2HDM have been proposed based on the couplings of the two Higgs doublets. Only the Type-III 2HDM possesses the generic Yukawa structure that includes the Higgs LFV effects and allows the *flavor-changing-neutral-currents* (FCNC) occur at tree-level [34]. Both φ_1 and φ_2 can couple to all fermions in this type. In type-I 2HDM, only φ_2 couples to charged fermions; in type-II, the up- and down-type quarks couple to separate doublets.

To derive the Higgs LFV couplings from 2HDM, it is necessary to rewrite the SM Lagrangian with two Higgs doublets. Start with equation 2.45, the Higgs term can be expressed as:

$$\mathcal{L}_{2HDM,Higgs} = (D_\mu \varphi_1)^\dagger (D^\mu \varphi_1) + (D_\mu \varphi_2)^\dagger (D^\mu \varphi_2) - V(\varphi_1^\dagger \varphi_1, \varphi_2^\dagger \varphi_2, \varphi_2^\dagger \varphi_1, \varphi_1^\dagger \varphi_2) \quad (3.26)$$

The general form of the scalar potential in the 2HDM is more complicated and includes more parameters than the SM one [34]:

$$\begin{aligned} V = & m_{11}^2 \varphi_1^\dagger \varphi_1 + m_{22}^2 \varphi_2^\dagger \varphi_2 - [m_{12}^2 \varphi_1^\dagger \varphi_2 + h.c.] \\ & + \frac{1}{2} \lambda_1 (\varphi_1^\dagger \varphi_1)^2 + \frac{1}{2} \lambda_2 (\varphi_2^\dagger \varphi_2)^2 + \lambda_3 (\varphi_1^\dagger \varphi_1) (\varphi_2^\dagger \varphi_2) + \lambda_4 (\varphi_1^\dagger \varphi_2) (\varphi_2^\dagger \varphi_1) \\ & + [\frac{1}{2} \lambda_5 (\varphi_1^\dagger \varphi_2)^2 + \lambda_6 (\varphi_1^\dagger \varphi_1) (\varphi_1^\dagger \varphi_2) + \lambda_7 (\varphi_2^\dagger \varphi_2) (\varphi_1^\dagger \varphi_2)] \end{aligned} \quad (3.27)$$

where m_{11} , m_{22} , λ_1 , λ_2 , λ_3 and λ_4 are real parameters, while m_{12} , λ_5 , λ_6 and λ_7 are complex parameters. The Higgs doublets can be decomposed as:

$$\varphi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{1}{\sqrt{2}}(v_1 e^{i\zeta_1} + \phi_1^0 + iC_1^0) \end{pmatrix}, \quad \varphi_2 = \begin{pmatrix} \phi_2^+ \\ \frac{1}{\sqrt{2}}(v_2 e^{i\zeta_2} + \phi_2^0 + iC_2^0) \end{pmatrix} \quad (3.28)$$

where v_1 (v_2) is the VEV of φ_1 (φ_2), ϕ^+ are the charged fields, ϕ^0 are the neutral CP-even fields and C^0 are the neutral CP-odd fields. Here the phases ζ also hint at the possibility of a broken CP-symmetry.

In type-III 2HDM, it is possible to change the bases without altering the physics by performing an arbitrary $U(2)$ transformation. Consider the doublets corresponding to mass eigenstates, where only one Higgs doublet acquires a non-zero VEV:

$$\varphi_1 = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{2}}(v + \phi_1^0 + iG^0) \end{pmatrix}, \quad \varphi_2 = \begin{pmatrix} H^+ \\ \frac{1}{\sqrt{2}}(\phi_2^0 + iA) \end{pmatrix} \quad (3.29)$$

where ϕ_1^+ and C_1^0 are redefined as the Goldstone bosons G , while ϕ_2^+ and C_2^0 are redefined as the

charged Higgs boson H^+ and CP-odd pseudoscalar A respectively. The mass eigenstates can be obtained by using the neutral scalars ϕ^0 and mixing angle θ :

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \sin \theta & \cos \theta \\ \cos \theta & -\sin \theta \end{pmatrix} \begin{pmatrix} \phi_1^0 \\ \phi_2^0 \end{pmatrix} \quad (3.30)$$

To derive the Yukawa couplings, one can start with the form of the LFV Lagrangian in equation 3.18, expressed in terms of the diagonal mass matrix $\hat{\mathcal{M}}$, as shown in equation 3.20. The leptonic part of the Yukawa Lagrangian in 2HDM is:

$$-\mathcal{L}_{2HDM,\ell}^Y = \frac{\sqrt{2}}{v} \bar{L} \hat{\mathcal{M}} \psi \varphi_1 + \bar{L} \rho \psi \varphi_2 \quad (3.31)$$

where ρ is the general matrix in flavour space. With equation 3.29 and equation 3.30, the Yukawa Lagrangian in $SU(2)$ space can be expanded as:

$$\begin{aligned} -\mathcal{L}_{2HDM,\ell}^Y &= \bar{\psi}_L \left[\frac{\hat{\mathcal{M}}}{v} \sin \theta + \frac{\rho}{\sqrt{2}} \cos \theta \right] \psi_R h \\ &+ \bar{\psi}_L \left[\frac{\hat{\mathcal{M}}}{v} \cos \theta - \frac{\rho}{\sqrt{2}} \sin \theta \right] \psi_R H \\ &+ \frac{i}{\sqrt{2}} \bar{\psi}_L \rho \psi_R A + \bar{\nu}_L (U^\dagger \rho) \psi_R H^+ + h.c. \end{aligned} \quad (3.32)$$

where U is the PMNS matrix in equation 3.8. The Yukawa couplings with the LFV effects of the neutral Higgs boson h can be extracted from the first bracket as:

$$Y_{h\ell\ell'}^{2HDM} = \frac{\hat{\mathcal{M}}_i}{v} \sin \theta \delta_{ij} + \frac{\rho^{ij}}{\sqrt{2}} \cos \theta \quad (3.33)$$

Which is in a similar form as equation 3.21, the first term is diagonal while the second term is the non-diagonal quantum fluctuations, the Higgs LFV effects in 2HDM could be coming from the 3×3 matrix ρ which can be expressed as [34]:

$$\rho^{ij} = \kappa_{ij} \frac{\rho^{\tau\tau}}{\sqrt{2}m_\tau/v} \sqrt{\frac{2m_i m_j}{v^2}} \quad (3.34)$$

where the contents of the κ_{ij} matrix are the parameters that correspond to different flavour indices. Additionally, as mentioned in section 3.1, the 2HDM can address the anomalous magnetic moment of muon, where $\tau\mu$ coupling with $\rho^{\tau\mu}$ in figure 3.1 can enhance the muon $g-2$ [40].

3.3.2 Supersymmetry

In the SM, fermions and bosons not only differ by half-integer in their spin quantum number but also behave very differently. However, supersymmetry introduces a symmetry between them: $Q_\alpha |j\rangle = |j \pm \frac{1}{2}\rangle$. In supersymmetry, every elementary particle has an associated particle known as *superpartner*, the spin of which differs by a half-integer. The bosonic partners of fermions could be the *squarks* and *sleptons*, while the fermionic partners of bosons could be the *gaugino* and the *Higgsino*. The superpartners share the same quantum number except for spin and have the same mass before the spontaneous symmetry breaking.

The supersymmetric framework that considers only the minimum number of new particle states and new interactions is known as the *Minimal Supersymmetric Standard Model* (MSSM). The MSSM

was originally proposed to solve the hierarchy problem discussed in section 3.1. In this model, the quadratic divergences in equation 3.2 can be cancelled out by the superpartners [54], as the couplings to the Higgs boson are the same:

$$\delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + \frac{\lambda_S}{16\pi^2}(2\Lambda^2) + \dots \quad (3.35)$$

where $\lambda_S = |\lambda_f|^2$ is the coupling of the superpartners to the Higgs boson. The contribution to the Higgs boson mass from radiative corrections is zero if superpartners are included.

In the Higgs sector, the MSSM has a two Higgs doublet structure in which one Higgs doublet couples to the up-type quarks and the other doublet couples to the down-type quarks, belonging to the type-II 2HDM. The 2HDM Lagrangian in equation 3.26 and the Higgs potential in equation 3.27 are applicable to the MSSM, but the parameters in the Higgs potential and the two Higgs doublets are modified with the features of supersymmetry. Considering two Higgs fields, H_1 with quantum number $Y = -1$ and H_2 with $Y = +1$, the Higgs potential can be derived by considering the quadratic term for $U(1)_Y$ and $SU(2)_L$, the superpotential term with $W = \mu H_1 H_2$ and the term for soft supersymmetry breaking ¹. Combining them in order gives [54]:

$$V_{MSSM} = \frac{g_B^2}{8}[(|H_2|^2 - |H_1|^2)]^2 + \frac{g_W^2}{8}[4|H_1^\dagger H_2|^2 - 2|H_1|^2|H_2|^2 + (|H_1|^2)^2 + (|H_2|^2)^2] \\ + \mu^2(|H_1|^2 + |H_2|^2) + [m_{H_1}^2 H_1^\dagger H_1 + m_{H_2}^2 H_2^\dagger H_2 + B\mu(H_1 H_2 + h.c.)] \quad (3.36)$$

The two Higgs doublets in MSSM can be expressed in terms of the charged and neutral components:

$$H_1 = \begin{pmatrix} \phi_1^{0+} \\ -\phi_1^- \end{pmatrix}, \quad H_2 = \begin{pmatrix} \phi_2^\dagger \\ \phi_2^0 \end{pmatrix} \quad (3.37)$$

By defining $\mu_1^2 = |\mu|^2 + m_{H_1}^2$, $\mu_2^2 = |\mu|^2 + m_{H_2}^2$, $\mu_3^2 = -B\mu$ and expanding the Higgs doublets, equation 3.36 can be re-derived as:

$$V_{MSSM} = \mu_1^2(|\phi_1^0|^2 + |\phi_1^-|^2) + \mu_2^2(|\phi_2^0|^2 + |\phi_2^\dagger|^2) - \mu_3^2(\phi_1^- \phi_2^\dagger + \phi_1^{0\dagger} \phi_2^0 + h.c.) \\ + \frac{g_W^2 + g_B^2}{8}(|\phi_1^0|^2 + |\phi_1^-|^2 - |\phi_2^0|^2 - |\phi_2^\dagger|^2)^2 + \frac{g_W^2}{2}|\phi_1^{0\dagger} \phi_2^\dagger - \phi_1^\dagger \phi_2^0|^2 \quad (3.38)$$

The spontaneous symmetry breaking happens if the conditions of $\mu_1^2 \mu_2^2 < \mu_3^4$ and $\mu_1^2 + \mu_2^2 > 2|\mu_3^2|$ are matched, and there are two VEVs in this model:

$$\langle H_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \sin \beta \end{pmatrix}, \quad \langle H_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v_2 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} v \cos \beta \\ 0 \end{pmatrix} \quad (3.39)$$

where $v^2 = v_1^2 + v_2^2$ and $\tan \beta = \frac{v_2}{v_1}$.

The contributions from the neutral Higgs LFV vertices can be derived by including the soft supersymmetry breaking as in equation 3.36, as there is no LFV at the superpotential level. The effective Lagrangian in this model can be written as [55]:

$$-\mathcal{L}_{eff, MSSM} = (2G_F^2)^{1/4} \frac{m_{\ell_i} \kappa_{ij}^\ell}{\cos^2 \beta} \bar{\psi}_{iR} [\cos(\alpha - \beta) h^0 + \sin(\alpha - \beta) H^0 - iA^0] \psi_{jL} + h.c. \quad (3.40)$$

¹A type of symmetry breaking that will not cause ultraviolet divergences to appear in the Higgs mass.

where $i \neq j$ and G_F is the Fermi constant, and κ_{ij}^ℓ is a matrix involving the mass terms and $\tan \beta$, the components of κ_{ij}^ℓ are introduced in [55, 56]. This effective Lagrangian can be used to calculate the LFV branching ratios in terms of the branching ratio of the SM flavour conserving decays $H \rightarrow \tau\tau$, using $H \rightarrow \mu\tau$ as an example [56]:

$$\mathcal{B}(\Phi^0 \rightarrow \mu\tau) = \tan^2 \beta (|\kappa_{\tau\mu}^\ell|^2) C_\Phi \mathcal{B}(\Phi^0 \rightarrow \tau\tau) \quad (3.41)$$

where $\Phi^0 = h^0, H^0, A^0$ and the coefficients C_Φ are given by:

$$C_h = \frac{\cos^2(\beta - \alpha)}{\sin^2 \alpha}, \quad C_H = \frac{\sin^2(\beta - \alpha)}{\cos^2 \alpha}, \quad C_A = 1 \quad (3.42)$$

The effective Lagrangian in equation 3.40 can also be used to estimate the other rare LFV processes such as $\tau \rightarrow 3\mu$ [55].

3.3.3 Composite Higgs Model

The Higgs mechanism was originally constructed based on the ideas of symmetry breaking in superconductivity, where spontaneously symmetry breaking is induced by the Cooper pair, a composite state with two electrons. Similarly, the quark-antiquark pair condensate induces the spontaneous breaking of a chiral symmetry in QCD. Therefore, it is reasonable to construct a model in which the Higgs boson is a composite particle.

The composite Higgs model with warped extra dimensions is one of the solutions to the hierarchy problem, while the model with warped extra dimensions is introduced explicitly in section 3.3.4. The *AdS/CFT correspondence* relates the composite Higgs model to the solution to the hierarchy problem based on warped extra dimensions. The basic idea is that physical fields can propagate through an extra dimension in a bulk piece, and the bulk behaves similarly to the strongly interacting sector. In this framework, the global symmetries are determined by those of the bulk. It is proposed in [57] that the $SO(5)$ symmetry group is broken via certain boundary conditions in a way: $SO(5) \rightarrow SO(4) \equiv SU(2)_L \times SU(2)_R \rightarrow SU(2)_L$, and the Higgs mass can be protected by the larger $SO(5)$ symmetry.

The general flavour structure of the Higgs boson couplings to fermions in the model with a warped extra dimension are introduced in [58]. The LFV contributions in this model could arise from the misalignment between these couplings and the fermion mass matrix. The explicit LFV contributions of the down-type quarks in the composite Higgs model are discussed in [59], while the author lefts the same generalizations in the charged lepton sector unexplained. In this section, the Higgs LFV effects from the charged lepton sector and in the Higgs composite model will be briefly discussed by following the same procedures in [59]. Begin with specifying the Wilson coefficients of the dimension-six operators in equation 3.14 and equation 3.16:

$$\begin{aligned} \mathcal{L}_e &= \frac{\kappa_{ij}^e}{\Lambda^2} i(\Phi^\dagger D^\mu \Phi)(\bar{\psi}_R \gamma_\mu \psi_{Rj}) + h.c. \\ \mathcal{L}_\ell &= \frac{\kappa_{ij}^\ell}{\Lambda^2} i(\Phi^\dagger D^\mu \Phi)(\bar{L}_i \gamma_\mu L_j) + h.c. \\ \mathcal{L}'_\ell &= \frac{\kappa_{ij}^{\prime\ell}}{\Lambda^2} i(\Phi^\dagger \tau^a D^\mu \Phi)(\bar{L}_i \gamma_\mu \tau^a L_j) + h.c. \\ \mathcal{L}_y &= \frac{\tilde{y}_{ij}^e}{\Lambda^2} (\Phi^\dagger \Phi)(\bar{L}_i \psi_{Rj} \Phi) + h.c. \end{aligned} \quad (3.43)$$

where e stands for the charged leptons. The field redefinition in term of $\tilde{y}_{ij}^e$:

$$\tilde{y}_{ij}^e \rightarrow \tilde{y}_{ij}^e + (\kappa^\ell y^\ell)_{ij} + (\kappa'^\ell y^e)_{ij} + (y^e \kappa^{e\dagger})_{ij} \quad (3.44)$$

where y^e is the Yukawa matrix. Similar to the $\hat{\lambda}'$ in equation 3.21, perform the same transformation to each of the Wilson coefficients:

$$\begin{aligned} y^d &\rightarrow V_L y^e V_R^\dagger, & \tilde{y}^e &\rightarrow V_L \tilde{y}^e V_R^\dagger, \\ \kappa^\ell &\rightarrow V_L \kappa^\ell V_L^\dagger, & \kappa'^\ell &\rightarrow V_L \kappa'^\ell V_L^\dagger, & \kappa^e &\rightarrow V_R \kappa^e V_R^\dagger \end{aligned} \quad (3.45)$$

Hence, the model-independent LFV term in equation 3.21 can be re-derived in terms of the coefficients as a form of non-diagonal Yukawa term: $\frac{M_i^e}{v} \delta_{ij} - \hat{y}_{ij}^e \frac{v^2}{\Lambda^2}$, where M^e is the lepton mass matrix and $\hat{y}_{ij}^e$ is the Higgs boson couplings similar to the $\mathcal{Y}$ in equation 3.19:

$$\begin{aligned} \frac{M_{ij}^e}{v} &= y_{ij}^e - (\tilde{y}^e + \kappa^\ell y^\ell + \kappa'^\ell y^e + y^e \kappa^{e\dagger})_{ij} \frac{v^2}{\Lambda^2} \\ \hat{y}_{ij}^e &= 2[D_L^\dagger \tilde{y}^e D_R]_{ij} - 2 \frac{M_j^e}{v} [D_L^\dagger (\kappa^\ell + \kappa'^\ell) D_L]_{ij} - 2 \frac{M_i^e}{v} [D_R^\dagger \kappa^{e\dagger} D_R]_{ij} \end{aligned} \quad (3.46)$$

where $D_{L,R}$ is the rotational matrices based on equation 3.45.

In the composite Higgs model, the strongly interacting sector involves bound composite states, including the Higgs doublet. For a 4D effective theory in 5D warped compactifications, the Higgs boson couplings to the SM fermions in this sector via the composite operators $O_{L,R}$ are assumed to be linear [59]:

$$\lambda_L \bar{\psi}_L O_R + \lambda_R \bar{\psi}_R O_L + h.c. \quad (3.47)$$

The Yukawa couplings can be expressed in term of the coupling of the composite Higgs boson to other strong states y_* [59]:

$$y^e \simeq y_* \frac{\lambda_L \lambda_R}{16\pi^2}, \quad \tilde{y}^e \simeq y_*^3 \frac{\lambda_L \lambda_R}{16\pi^2}, \quad \kappa^\ell, \kappa'^\ell \simeq y_*^2 \frac{\lambda_L^2}{16\pi^2}, \quad \kappa^e \simeq y_*^2 \frac{\lambda_R^2}{16\pi^2} \quad (3.48)$$

where a_{ij} is a matrix with $0(1)$ components. By putting the coefficients back in equation 3.46 and assuming the D_L has similar components as the PMNS matrix, one can write down the Higgs couplings to a pair of leptons with different flavours as $\hat{y}_{ij}^e \sim y_*^2 \frac{M_j}{v}$. Therefore, a misalignment between the Higgs couplings and the lepton mass matrix is illustrated in this relation, which could lead to non-zero LFV contributions.

3.3.4 Randall-Sundrum Model

The *Randall-Sundrum model* is a theory with 5-dimensional warped geometry that was developed to solve the hierarchy problem. In this model, the universe is described with a warped geometry with higher-dimension. The elementary particles are localised on a (3+1)-dimensional brane in a 5-dimensional *anti-de Sitter space*. Unlike De Sitter space, which corresponds to gravity, anti-de Sitter space corresponds to the fundamental forces in the Standard Model, and this is known as the *AdS/CFT correspondence*.

The Randall-Sundrum model provides a solution to the hierarchy problem between the electroweak scale and Planck scale in a higher-dimensional scenario. In this model, the elementary particles and forces in the SM are confined to the "3-brane", which is the 4-dimensional subspace

within the 5-dimensional spacetime. To parametrize the fifth dimension, the phase ϕ with values $0 \leq \phi \leq \pi$ is used. The boundaries of the 5-dimensional spacetime are defined by the two 3-branes located at $\phi = 0$ and $\phi = \pi$, extending in the x^μ direction. The SM fields are localized on the brane with $\phi = \pi$. Moreover, the solutions to the 5-dimensional Einstein's equations are given in [60], which has the form:

$$ds^2 = e^{-2kr_c\phi} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2 \quad (3.49)$$

where r_c is known as the "compactification radius", $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$ and the scale of k is in the same order of Planck scale. The exponential factor accounts for the warped geometry of the extra dimensions. The authors in [60] also show that the physical mass scales in this model are defined by a symmetry-breaking scale $v \equiv e^{-kr_c\pi} v_0$. In other words, the mass scale m_0 on the 3-brane with $\phi = \pi$ is proportional to the physical mass through the exponential term: $m \equiv e^{-kr_c\pi} m_0$. In the case $m_0 = M_{Planck}$, the hierarchy problem is solved if $e^{-kr_c\pi} \sim 10^{15}$. The hierarchy between the 4-dimensional masses on the brane and the Planck mass can be explained by the warped extra-dimension or the exponential term.

The Higgs LFV contributions arise in the 5-dimensional warped geometry theory could be due to the misalignment between the Higgs Yukawa matrices and the fermion mass matrices, similar effects are expected in the composite Higgs model and discussed in section 3.3.3. The LFV decays of the Higgs boson has been studied in [61], the authors have addressed the LFV Higgs couplings in both the brane Higgs limit and bulk Higgs limit, and presented the Higgs LFV Lagrangian:

$$\mathcal{L}_{HLFV} = a_{ij} \sqrt{\frac{m_i m_j}{v^2}} H \bar{\psi}_L^i e^j + h.c. \quad (3.50)$$

where $m_{i,j}$ is the lepton mass and e stands for the charged leptons. The a_{ij} is a non-diagonal matrix whose components depends on the profiles $f_{L,e}$ and could lead to the Higgs LFV contributions, while its expressions and numerical results are introduced in [61].

3.3.5 Seesaw Mechanism

The observation of neutrino oscillations, as discussed in section 3.1.1, indicated that neutrinos have mass, the SM must be extended to explain the extremely small neutrino masses. One simple extension is the inclusion of right-handed neutrinos, which are absent in the SM.

Neutrinos in the SM are *Dirac particles*, meaning their antiparticles are distinct. However, *Majorana particles* are fermions that are their own antiparticles. The property of having zero electric charge allows neutrinos to have the Majorana masses. The mass matrix of the neutrinos and its eigenvalues are:

$$M_\nu = \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix}, \quad m_\pm = \frac{1}{2} [(m_R + m_L) \pm \sqrt{(m_R - m_L)^2 + 4m_D^2}] \quad (3.51)$$

where m_L (m_R) is the left-handed (right-handed) Majorana mass term. The left-handed neutrinos are protected by the $U(1)_Y$ symmetry, one can define $m_L = 0$ as the Majorana mass term with left-handed neutrinos induce a change in the hypercharge after electroweak symmetry breaking. Processes involving left-handed Majorana mass term would violate the lepton number by 2. Thus, the extension with a right-handed $U(1)$ singlet Majorana neutrino field is motivated, and this extension is

known as the type-I *seesaw mechanism*. The heavy Majorana neutrinos with $m_R \gg m_D$ are expected in this model, and the eigenvalues in equation 3.51 correspond to a light left-handed neutrino of mass m_D^2/m_R and a heavy right-handed neutrino of mass m_R . By solving the five-dimensional effective Lagrangian with the Weinberg operator after the symmetry breaking, a Majorana neutrino mass $Y_\nu^T \frac{1}{m_R} Y_\nu v^2$ can be obtained.

Type-I seesaw is the simplest type of seesaw mechanism, but there are other different types such as the common type-II and type-III seesaw. The type-II seesaw introduces one new scalar triplet $\Delta = (\Delta_1, \Delta_2, \Delta_3)$, and the type-III seesaw introduces a $SU(2)$ triplet Σ . The matrix form of Δ and Σ are:

$$\Delta = \begin{pmatrix} \Delta^+/\sqrt{2} & \Delta^{++} \\ \Delta^0 & -\Delta^+/\sqrt{2} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_0/\sqrt{2} & \Sigma_+ \\ \Sigma_- & -\Sigma_0/\sqrt{2} \end{pmatrix} \quad (3.52)$$

where Δ^+ , Δ^0 and Δ^{++} are derived by Pauli matrices and in terms of $\Delta_{1,2,3}$, Σ_+ and Σ_- form a Dirac fermion and Σ_0 is a Majorana fermion. After the symmetry breaking, the Δ^0 in type-II seesaw obtains a VEV $v_\Delta = \frac{\mu v^2}{\sqrt{2} M_\Delta^2}$, where M_Δ is the mass in the field Δ and μ is a coupling constant. Therefore, the neutrinos acquire a Majorana mass $Y_\Delta \frac{\mu}{M_\Delta^2} v^2$. While the generation of neutrino mass in type-III seesaw is similar to the mechanism in type-I, with the Weinberg operator in type-III, the neutrino acquires a Majorana mass $Y_\Sigma^T \frac{1}{M_\Sigma} Y_\Sigma v^2$. These types of seesaw mechanism can generate the five-dimensional operator at tree level, which generates the Majorana neutrino masses after the electroweak symmetry breaking, explaining the neutrino mass generation in an elegant way.

Considering that new physics models need to explain neutrino mass and that the observations of neutrino oscillations are the only evidence of LFV, it is natural to expect charged LFV effects to arise in the seesaw mechanisms that explain neutrino mass generation. There are non-zero LFV contributions at the loop level in different types of seesaw mechanisms. By including the right-handed neutrinos, the Yukawa coupling matrix is not necessarily diagonal.

The LFV decay rates are much larger in the *inverse-seesaw mechanism* compared to the other types. The inverse-seesaw (ISS) allows for large Yukawa neutrino couplings, while small neutrino masses arise as new physics at the TeV scale. The ISS introduces three pairs of additional fermionic singlets (ν_R, X) with opposite lepton number. The Lagrangian of ISS can be written as [62]:

$$\mathcal{L}_{ISS} = -Y_\nu^{ij} \bar{L}_i \tilde{\Phi} \nu_{Rj} - M_R^{ij} \bar{\nu}_{Ri}^C X_j - \frac{1}{2} \mu_X^{ij} \bar{X}_i^C X_j + h.c. \quad (3.53)$$

where $\tilde{\Phi} = i\sigma_2 \Phi^*$ is the SM Higgs doublet, M_R and μ_X are the 3×3 matrices. After the electroweak symmetry breaking, the mass matrix in the ISS can be obtained as:

$$M_{ISS} = \begin{pmatrix} 0 & m_D & 0 \\ m_D^T & 0 & M_R \\ 0 & M_R^T & \mu_X \end{pmatrix} \quad (3.54)$$

In contrast to the type-I seesaw (inverse of it), consider $\mu_X \ll m_D \ll M_R$ such that the 9×9 mass matrix can provide an effective neutrino mass matrix of the form $m_\nu = m_D^T (M_R^T)^{-1} \mu_X M_R^{-1} m_D$. Which can be diagonalized by using the PMNS matrix, and one can retrieve the lepton number symmetry by setting the μ_X to zero.

The details of the Higgs LFV contributions at one loop level are given in [62], in specific, figure 1 in this paper shows all the contributing one-loop Feynman diagrams. Define that the amplitudes F_L and F_R are the summation of all the one-loop contributions, the decay width of $\Gamma(H \rightarrow \ell_i \bar{\ell}_j)$ can be obtained as:

$$\Gamma(H \rightarrow \ell_i \bar{\ell}_j) = \frac{g^2}{16\pi m_H} \sqrt{\left[1 - \left(\frac{m_{\ell_i} + m_{\ell_j}}{m_H}\right)^2\right] \left[1 - \left(\frac{m_{\ell_i} - m_{\ell_j}}{m_H}\right)^2\right]} \times [(m_H^2 - m_{\ell_i}^2 - m_{\ell_j}^2)(|F_L|^2 + |F_R|^2) - 4m_{\ell_i}m_{\ell_j} \text{Re}(F_L F_R^*)] \quad (3.55)$$

The branching ratio of other LFV decays such as $\ell_i \rightarrow \ell_j \gamma$ are also computed in [62] and the values are within the experimental upper bounds.

Other seesaw mechanism such as the MSSM-seesaw also predicted the large branching fractions of the Higgs LFV decays [63]. The SM-like seesaw mechanism predicts much smaller LFV branching fractions due to the decoupling behaviour of the heavy Majorana neutrinos and smallness of the neutrino masses. In [63], the authors found that the SUSY-like seesaw mechanism predicts significantly larger LFV decay rates.

Chapter 4

The LHC and the ATLAS Detector

The primary goal of modern experimental particle physics is to validate the SM with higher precision measurements and probe new physics at higher energy. Both can be achieved in the particle accelerators, which can generate a sufficiently high collision rate and high energy. A huge machine will be needed for accelerating the particle beams to sufficiently high energy due to the synchrotron radiation. Luckily, the world's largest and most complex machine - the Large Hadron Collider - was granted by the CERN council in 1994 [64].

The second operation (Run-2) of LHC can generate collisions with centre-of-mass energy $\sqrt{s}=13$ TeV, which is on a large scale where new physics may arise at the LHC. On the other hand, the search of those rare decays with very low branching ratio demands extremely high luminosity. In the LHC, the number of events per second is given by $N = L\sigma_{event}$, where L is the luminosity and σ_{event} is the cross section of an event. During Run-2, there are almost 8 million Higgs bosons have been produced, making the LHC the best high-rate source for searching for the signals of the objective Higgs boson decays. In the third operation (Run-3) started last year, the peak integrated luminosity is expected to reach 450 fb^{-1} with $\sqrt{s} = 13.6$ TeV, and 18 million of the Higgs bosons are expected to be produced. The LHC has become a Higgs boson factory in the world and will extend the frontier of particle physics.

In this section, the overview of the LHC and the ATLAS detector is introduced. The acceleration of the particle beams in the LHC and their collision processes in the ATLAS detector are also discussed. During the second long shutdown period, the ATLAS detector has been upgraded in order to adapt the large increase in the luminosity and energy in Run-3, these upgrades are introduced.

4.1 The Large Hadron Collider

The Large Hadron Collider, located at CERN, is the largest and highest-energy particle accelerator in the world, it is built in a tunnel with a circumference of 27 kilometers at a depth of 175 meters underground [64], which provides shielding from background radiations. Inside the ring, the particles beams are guided by electromagnetic fields induced by the superconducting magnets, and the magnets are cooled to a temperature close to absolute zero using the liquid helium system to ensure efficient operation. As the electromagnetic fields increase from 0.54 to 7.7 T, two proton beams can be accelerated from 450 GeV to 6.5 TeV in the opposite direction, reaching a speed of 99.999999% of the speed of light before colliding.

The first successful collisions between two proton beams with energy of 3.5 TeV were achieved

in 2010, the energy of particle accelerator have come into a new era since then. Just before the end of the runs in the first operation in 2012, both ATLAS and CMS detectors observed signals of new particle - the Higgs boson - in the data obtained.

The second operation started in 2015 after the two years long shutdown (LS1). During the LS1 period, both the accelerator and the detectors underwent upgrades to enable the 14 TeV proton-proton collisions. However, the proton beams were only accelerated to 6.5 TeV, resulting in recorded collisions of up to 13 TeV, as the energy scale was limited by the capability of the bending magnets. After three years operation from 2015 to 2018, the integrated luminosity of the full Run-2 dataset is measured as $140 \pm 1.2 \text{ fb}^{-1}$ [65].

There are large upgrades in the second long shutdown period for enabling the particle beam energy of 13.6 TeV and much higher luminosity, the upgrades on the ATLAS detector are discussed in section 4.2 specifically. The third operation started last year on April 2023, the total luminosity is expected to double compared to the Run-2 dataset. The LHC has been turned on again for another three years operations, scheduled to continue until the end of 2025, with the expectation of further advancing the frontier of particle physics through the acquisition of a much larger dataset.

The integrated luminosity in Run-3 is expected to reach 450 fb^{-1} , setting a new record in particle accelerators. However, the LHC is expected to perform even better in the High-luminosity-LHC (HL-LHC) projectm which will take place after Run-3 and LS3. The integrated luminosity is expected to reach 3000 fb^{-1} . Additionally, the expected HL-LHC performance of the objective analysis is studied by the extrapolation procedures which are discussed in section 15.

For more detailed information on the LHC, the first LHC design report [64] provides comprehensive details. In this section, an overview of the LHC based on this report is provided, along with additional information on the upgrades implemented during the long shutdown periods.

4.1.1 Layout

The LHC circle is divided into eight arcs, each serving different functions and detectors. The schematic layout of the LHC is shown in figure 4.1. Two particle beams are injected in opposite direction from octant 2 and octant 8, represented by the red and blue curves in figure 4.1. Each arc in the ring consists of arc cells, and multiple multipole magnets are installed in each arc cell. Overall, there are 1232 dipole magnets responsible for bending the particle beams in the circle, and 392 quadrupole magnets for keeping them focused [64]. During the LS2 period, the dipole magnets were upgraded with new insulating caps to improve the electrical insulation.

The particle beams intersect at the four detectors: the A Toroidal LHC ApparatuS (ATLAS) detector in octant 1, A Large Ion Collider Experiment (ALICE) in octant 2, the Compact Muon Solenoid (CMS) in octant 5, and the LHC-B experiment for studying physics related to B-meson decays in octant 8. ATLAS and CMS are general detectors focusing on studying the Higgs boson and searching for signals of new physics, they have similar subdetectors with their own advantages but differ in the magnetic field strengths and calorimeters. On the other hand, ALICE and LHC-B are special detectors that have different main purposes. ALICE is optimized for studying the quark-gluon plasma by detecting the outputs from the heavy-ion collisions, which allows to re-generate a very similar condition just after the Big Bang. While the LHC-B experiment is specialized for b-physics, unlike the ATLAS and CMS detector which covers almost entire solid angle around the collision point, the LHC-B focus on detecting the forward particles. The main goals of LHC-B are studying the matter-antimatter asymmetry and CP violation. In all these four detectors, the particle collisions

happen in the straight sections as illustrated by stars in figure 4.1.

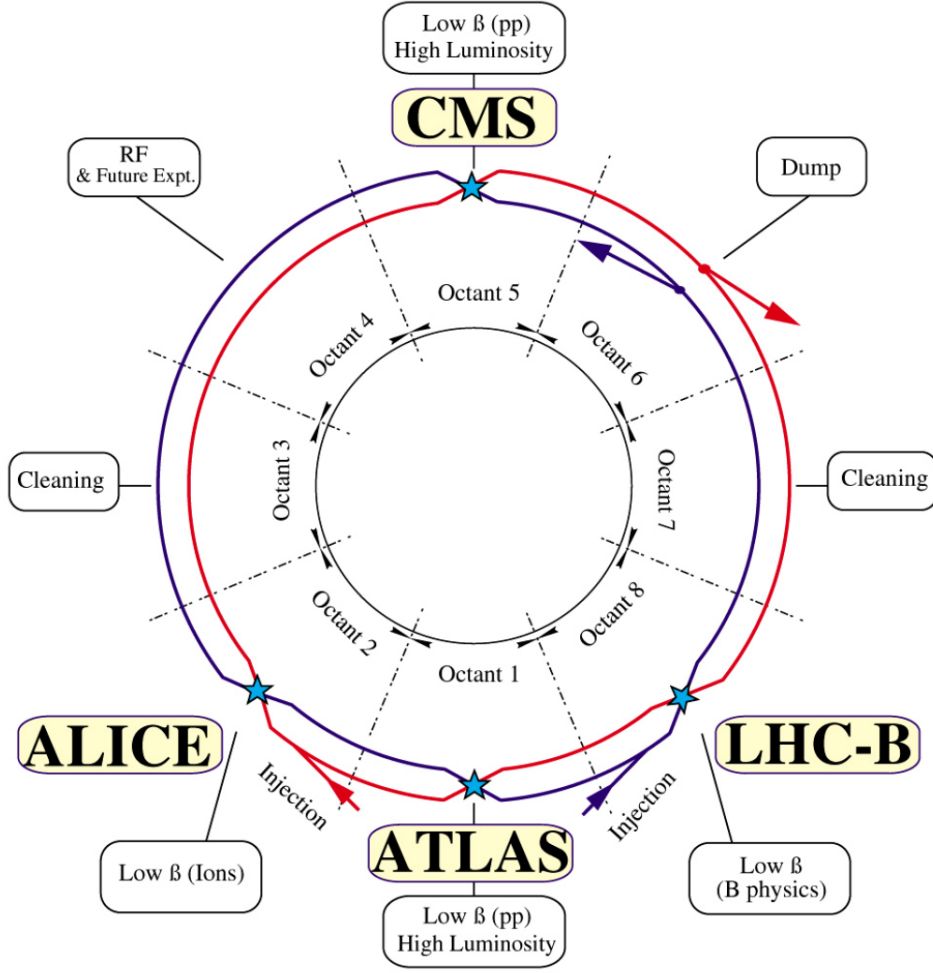


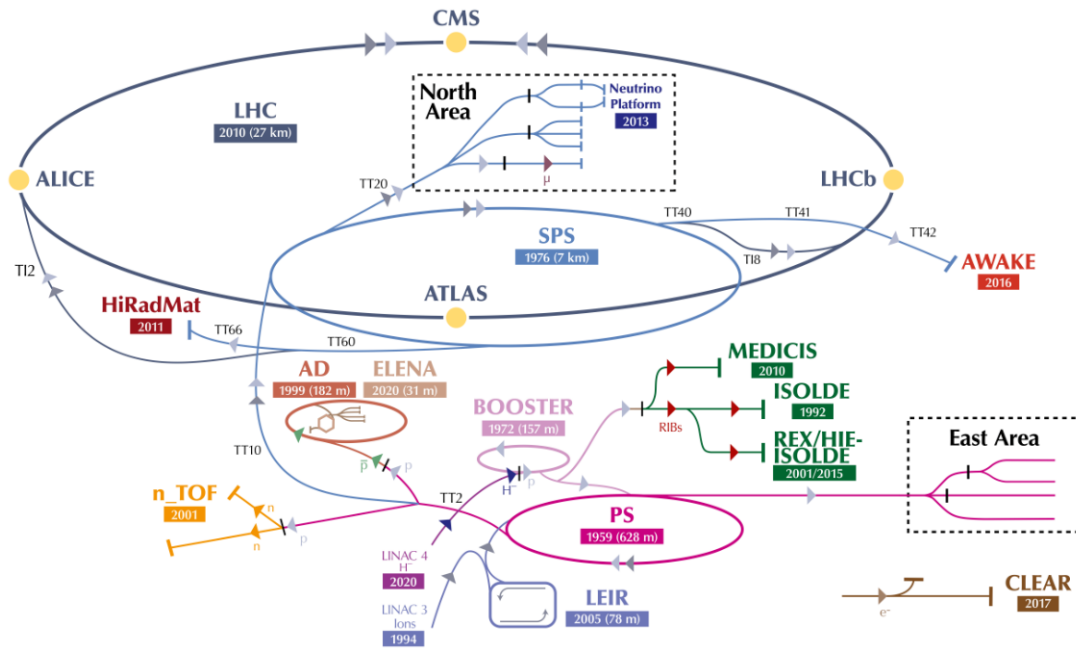
Figure 4.1: The schematic diagram of the LHC layout [66]

The LHC must be protected from unavoidable beam losses, the intensity and energy of the proton beams are extremely high such that a small fraction of the stored beams is able to quench the superconducting magnets or even damage the accelerator. Therefore, those strayed protons - or so called "beam halo" must be cleaned or absorbed for avoiding beam induced quenches. The LHC collimation system provides the cleaning in two main cleaning insertions, the one in octant 3 is responsible for the momentum cleaning and another in octant 7 is responsible for the betatron cleaning. During the LS2 period, 16 new collimators were installed to accommodate the high-energy particle beams.

The Radio Frequency (RF) cavities located in octant 4 are responsible for the particle acceleration. The injected beams are accelerated by a 400 MHz superconducting cavity system [64]. When the charged particles enter the RF cavities, they receive electrical pulses and hence being accelerated. The cryogenic power was increased during the LS2, which is a crucial upgrade for the HL-LHC project.

The LHC beam dump in the octant 6 is designed to make a fast extraction of the beams in a loss-free way, they are located far from the ring such that the absorber will not be overheated by the beams. The absorber materials are made of graphite and blocked by a cylinder, which should be

able to accept the beam energy. The dump is also required to absorb the beams when it is time to get rid of the beams or even in an emergency, the beams will be kicked to the dump by the kicker magnets in these cases. During the LS2, the two original beam dumps have come to the end and they were replaced by the spare dumps which are much more powerful and stable to meet the requirements of the HL-LHC project.



The LHC accelerates the injected proton beams from 450 GeV to 6.5 TeV, but before the injections, a chain of accelerators are responsible for increasing the energy of protons up to 450 GeV. Figure 4.2 shows the schematic diagram of the accelerator chain at CERN, the complex system feeds the particles from one accelerator to the next, cooperating to ramp up the energy of the particles gradually.

- **Linear accelerator 4 (Linac4):** A linear accelerator that uses the RF cavities to charge the cylindrical conductors, and the conductors accelerate the charged particles. Linac4 is the source of the proton beams in the chain, it accelerates the negative hydrogen ions to 160 MeV before injecting them to the PSB. During the injection process from Linac4 to PSB, the ions abundant their two electrons and only the protons were injected into PSB.
- **Proton Synchrotron Booster (PSB):** The PSB containing four superimposed synchrotron rings, which further accelerates the 160 MeV protons from the Linac4 to 2 GeV, and those 2 GeV protons are injected into the PS.

- **Proton Synchrotron (PS):** The circle in the PS has a large radius of 14.1 m, including 100 dipole magnets for guiding the beam around the ring. The PS will accelerate the injected 2 GeV protons to 26 GeV, in addition, it also can be used to accelerate the heavy ions from the Low Energy Ion Ring (LEIR). The PS has a long history since 1959 [68], it was used to accelerated many other types of particle for many experimental purposes.
- **Super Proton Synchrotron (SPS):** The 26 GeV protons from the PS will be injected into a bigger circle - the SPS for achieving energy up to 450 GeV. The SPS is the second-largest machine at CERN, with 744 dipole magnets for bending the beams around the ring [69]. In addition to inject the 450 GeV particles into the LHC, the beams come from SPS will also be fed into the Advanced Wakefield Experiment (AWAKE) for developing the new techniques of plasma acceleration, and the beams will be fed into the neutrino experiments as well.

The accelerator complex system not only serve the LHC, but also contributes to many other experiments at low energies, exploring many other areas such as the neutrinos, antimatters, dark matter and many others.

4.2 The ATLAS Detector

ATLAS is the largest general-purpose detector at CERN. It is designed to measure a wide range of signals instead of any particular process. The ATLAS detector is cylindrical in shape, with a length of 44 meters and a diameter of 25 meters [70]. It is located at Meyrin, Geneva, just near the main campus of CERN. The extremely high proton-proton collision rate and energy are significant challenges to the design of the ATLAS detector, the variety types of physical objects from the collisions require different subsystems (or subdetectors) which are designed according to the features of different objects.

In this section, the details of the ATLAS detector are introduced based on [70]. In the LS2 period, there are important upgrades on the ATLAS detector for enabling the increase in the energy and luminosity in the Run-3 experiments and beyond, the details of the LS2 upgrades are also discussed.

4.2.1 Layout

The ATLAS detector is expected to be able to detect any new physical processes if they happened after the collisions, and it should also be able to measure the properties of the physical objects such as their masses, energies, spins, lifetime and charges. The cutaway view of the ATLAS detector is shown in figure 4.3, and it exhibits a forward-backward symmetry with respect to the interaction point.

The detector consists of four main subsystems, each of them composed multiple layers. The inner detector is closest to the proton beams and is surrounded by the solenoid magnet. Outside the magnet are calorimeters, and there are two calorimeter systems - the electromagnetic calorimeter and hadronic calorimeter. The muon spectrometer is an extremely large tracking system which forms the outer part of the detector. While the toroid magnet and solenoid magnet are used to bend the charged particles so that their momentum can be measured. The ATLAS detector is designed to cover almost the entire solid angle around the collision point.

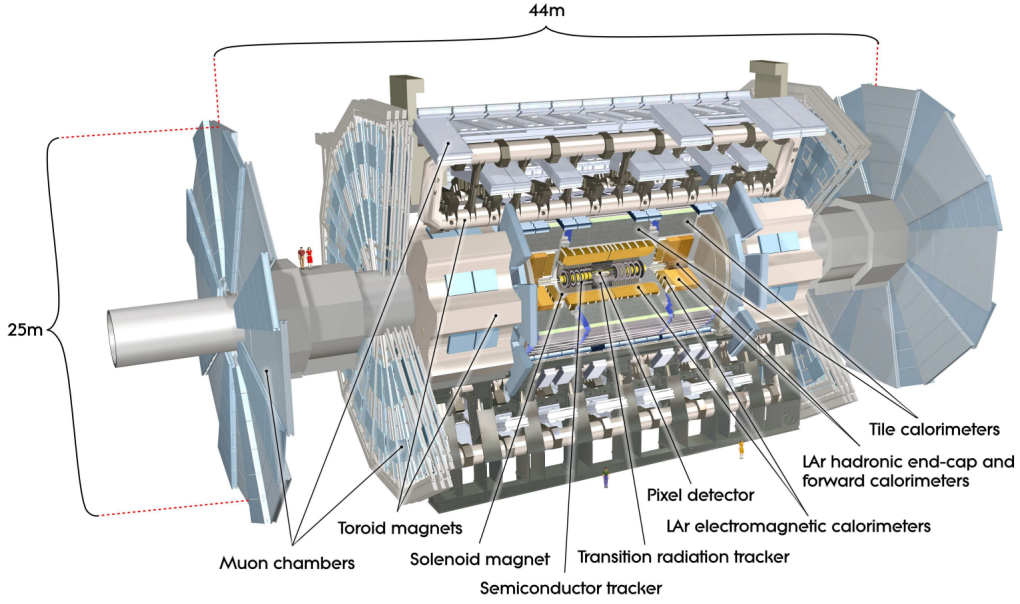


Figure 4.3: The longitudinal cutaway layout of the ATLAS detector [70]

Each of the main subsystem is designed focusing on different features and functionalities according to the properties of the physical objects, the details are discussed in this chapter.

4.2.2 Coordinate System

ATLAS experiments utilize a right-handed coordinate system, with the interaction point defined as the origin of the coordinate system. The positive x-axis is pointing from the interaction point to the centre of LHC ring, while the positive y-axis is pointing upward, and the z-axis is the beam direction. The angle around the beam axis is the azimuthal angle ϕ , while the polar angle θ is measured from the beam axis. The structure of the coordinate system is shown in figure 4.4.

In hadron collider physics, it is more common to use the *pseudorapidity* η to represent the angle between the particle's trajectory and the beam axis in the form that:

$$\eta = -\ln\left(\tan \frac{\theta}{2}\right) \quad (4.1)$$

Which means the particles fly in the direction which is perpendicular to the beam axis have $\eta = 0$, and those particles with very large η usually escape the detector and lost along the beam. Pseudorapidity is primarily used for the massless particles, while another quantity called *rapidity* is derived for the massive particles:

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) \quad (4.2)$$

where E is the energy of the particles and p_z is the momentum along the z-axis. The $\Delta\eta$ and Δy between particles are Lorentz invariant quantities and independent of the longitudinal boosts of the reference frame, therefore, they are much more popular than the polar angle θ in the high energy experiments. Moreover, the angular distance between particles which is Lorentz invariant is also very useful and can be defined as:

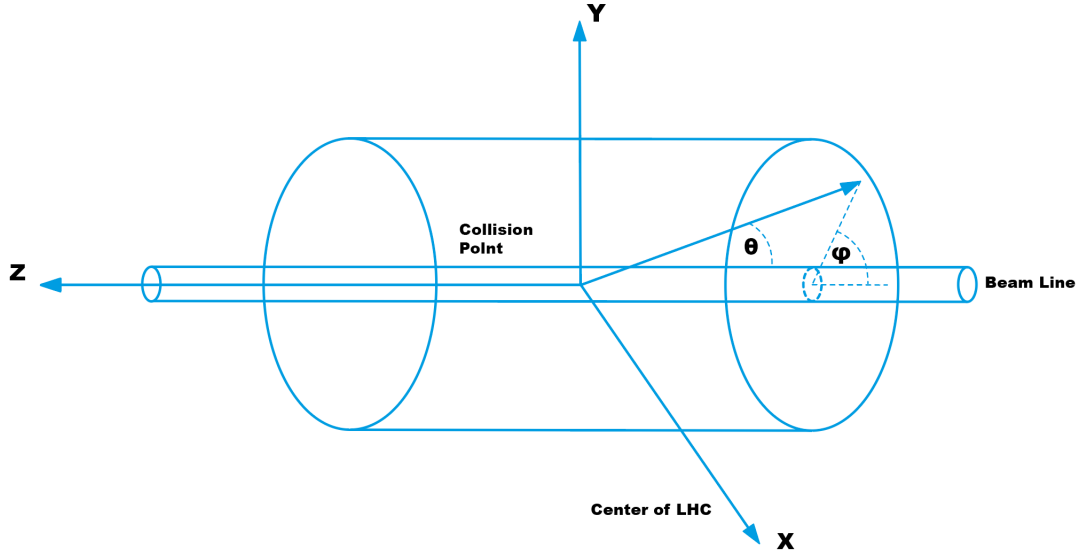


Figure 4.4: The coordinate system used by the ATLAS experiments.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (4.3)$$

In experimental particle physics, the transverse momenta of particles are defined in Cartesian coordinate, measured in term of the transverse plane ϕ and pseudorapidity η :

$$p_x = p_T \cos\phi, \quad p_y = p_T \sin\phi, \quad p_z = p_T \sinh\eta \quad (4.4)$$

The transverse mass, which is Lorentz invariant under boosts along the beam axis can be defined in term of p_T :

$$m_T = \sqrt{p_T^2 + m^2} \quad (4.5)$$

4.2.3 Inner Detector

Figure 4.5 shows the path of different particles in the ATLAS detector, the inner detector is the first part of the ATLAS detector that meets the decay products from the collisions. It is immersed in a 2 T magnetic field provided by the solenoid magnets [70]. While the magnets bend the charged particles, the direction and the degree of curvatures of the charged particles are recorded by the inner detector so that their charge and momentum can be measured. The inner detector consists of three main components: the pixel detector, the semiconductor tracker and the transition radiation tracker, a cutaway view of the inner detector is shown in figure 4.6. The inner detector is designed to provide precision reconstruction of the tracks of charged particles, primary and secondary vertex from the collisions, and sufficiently high reconstruction efficiency.

The large track density at the LHC requires the inner detector to have a sufficiently high granularity for providing the high momentum and vertex resolution. It also covers nearly entire

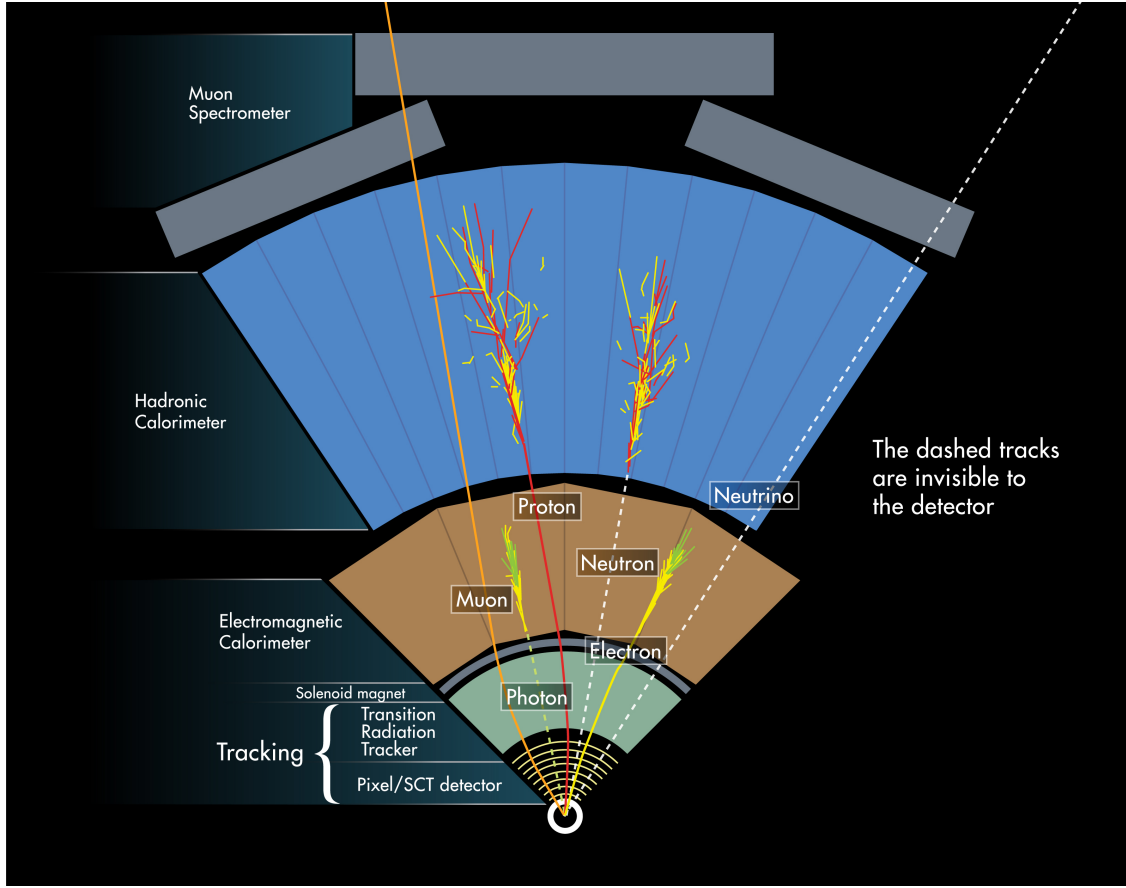


Figure 4.5: The schematic view of the particle path in the ATLAS detector [71].

azimuthal angle and has a large acceptance in pseudorapidity. Figure 4.7 shows a plane view of a quarter-section of the inner detector, the pixel detector, which is the precision tracking detector, and the semiconductor tracker in the end-cap region have covered the region with $|\eta| < 2.5$, and the transition radiation tracker also provides high quality electron identification and high momentum resolution in the region $|\eta| < 2.0$.

The pixel detector is the innermost part of the inner detector, it is located only 3.3 cm from the beam line. It comprises four layers of silicon pixels, totaling 92 billion pixels and nearly 2000 elements in the detector. This compact design allows for extremely high precision tracking of charged particles. The charged charges can be tracked by using the energy deposits or hits in the pixels and other detectors, which cooperates to reconstruct the trajectory of the charged particles.

Surrounding the pixel detector is the semiconductor tracker, located in the middle part of the inner detector. It is used to reconstruct the tracks of the charged particles by using six million readout strips which are long and narrow in size, with a larger area of silicon compare to the pixels and distributed over four double layers. The large coverage in area provides an extremely high precision measurements on the tracks of charged particles, with a similar accuracy compare to the pixel detector.

The outermost part of inner detector is the transition radiation tracker, the detecting elements are 300,000 drift tubes (straws). Each straw is filled with gas that becomes ionized when the charged particles pass through, and hence the induced electric pulse that can be used to reconstruct the tracks of charged particles with high precision. The transition radiation tracker is also used to

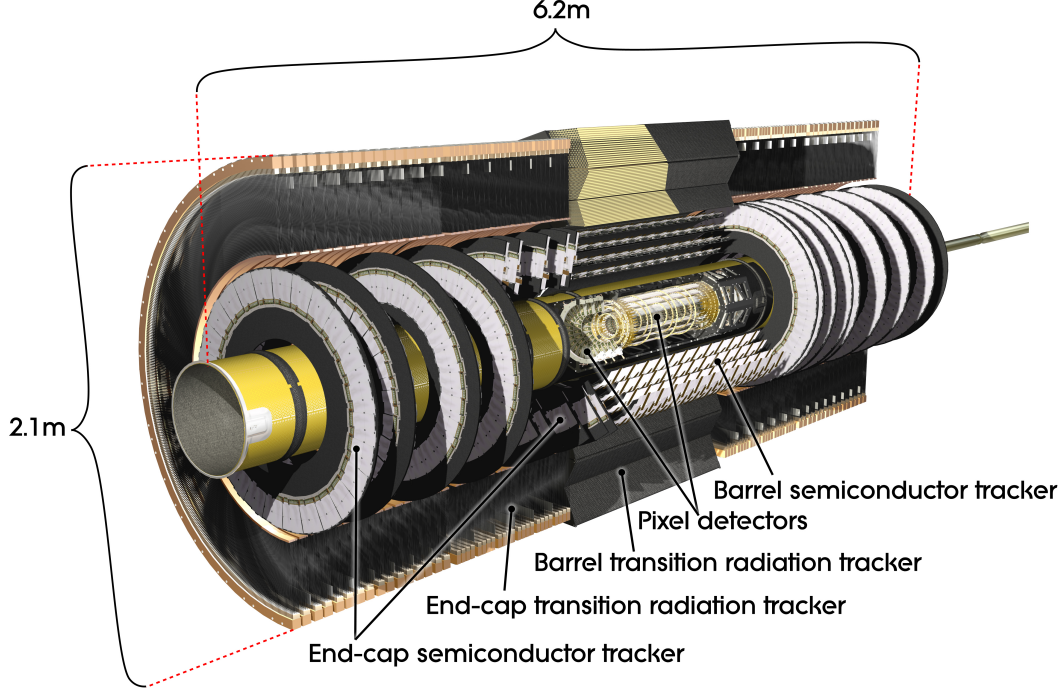


Figure 4.6: The cutaway layout of the ATLAS inner detector [72].

discriminate between electrons and hadrons when the charged particles cross through the detector.

4.2.4 Calorimeters

The ATLAS calorimeters are installed outside the solenoid magnets, there are two types of calorimeters: the Liquid argon (LAr) electromagnetic calorimeter and the tile hadronic calorimeter. The former is located in the inner part, while the latter is located in the outer part. The cutaway view of the calorimeters is shown in figure 4.4.

The calorimeters are designed to measure the energy of particles by absorbing their energies and stopping them. Most particles can be stopped in the detector except the muons and neutrinos. As shown in figure 4.5, muons can be detected by the muon spectrometer, while neutrinos will penetrate the ATLAS detector directly and being accounted in the missing energy. The calorimeter system has a large coverage in pseudorapidity up to $|\eta| < 4.9$, the forward region with large pseudorapidity is covered by the LAr forward calorimeter.

The ATLAS calorimeters are also known as sampling calorimeters. When the particles are being stopped and interact with the detecting elements, they induce showers of energy in the calorimeters, and the shape of the particle showers is sampled periodically. Only a fraction of the particle energy is measured, while the full energy can be inferred by using the particle showers.

4.2.4.1 Electromagnetic Calorimeter

The LAr electromagnetic calorimeters in the barrel and endcap regions are designed to measure the absorbed energies of the electrons and photons, the location of the energy deposits can also be provided in high precision. The electromagnetic calorimeter covers a pseudorapidity range up to $|\eta| < 3.2$, the electrodes in accordion shape can cover the full azimuthal angle without cracks. The

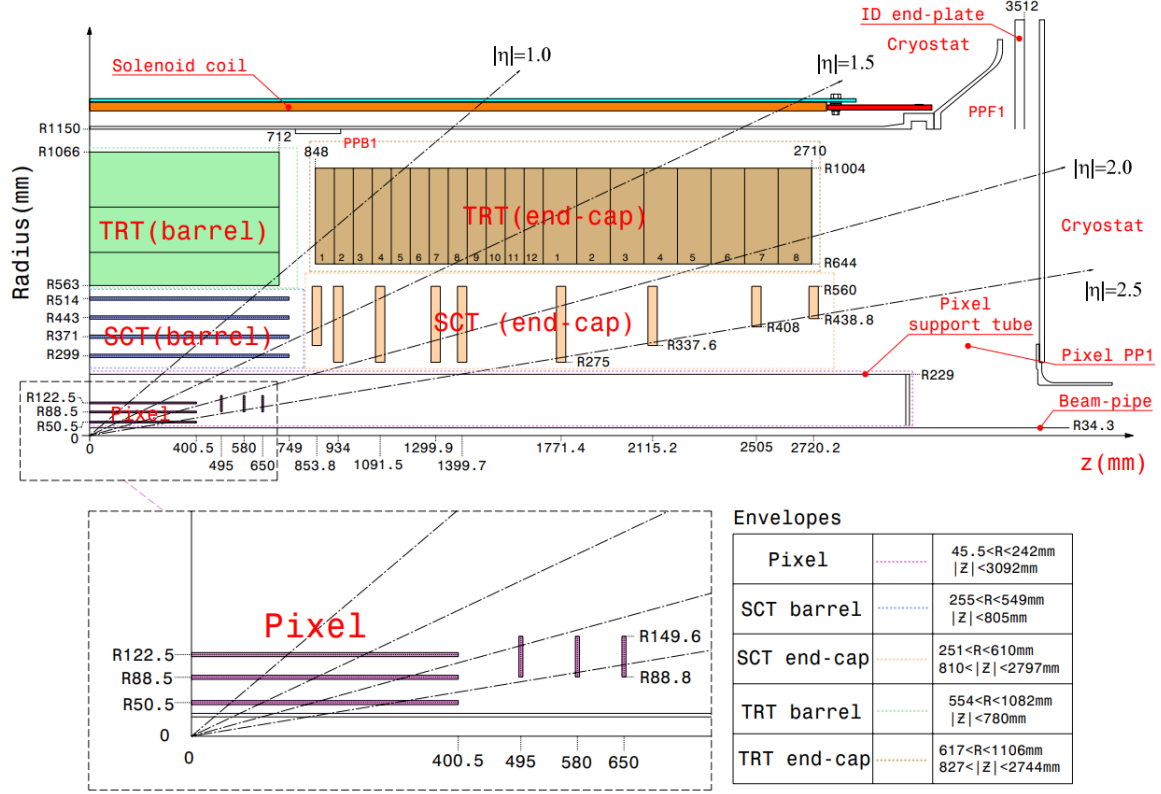


Figure 4.7: The plane view of a quarter-section of the ATLAS inner detector, the envelopes of each of the main component are also shown [70].

materials used for energy absorbing are the lead and stainless steel, while the liquid argon is used as the sampling material. To maintain stability and performance of the electromagnetic calorimeter and keeping the argon in liquid state, the cryostat is installed around the calorimeter for keeping it in a low enough temperature.

When charged particles pass through, the absorbing materials in three different layers absorbed the energy of the particles, converting them into showers, figure 4.9 shows the three layers in the barrel region. The liquid argon are ionized by the particles, inducing the electric current that will be measured. The position of the particles can be measured in very high precision by the first layer, and the identification of the particles are provided by the central part with full azimuthal angle.

The geometry of the barrel electromagnetic calorimeter is shown in figure 4.9 and figure 4.8, there are two parts of barrels around the beam axis. They have covered the pseudorapidity range of $|\eta| < 1.475$, one of them cover the positive η region and the other cover the negative η region. The trigger tower has a size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, the energy deposited in the trigger towers are provided to the first level (L1) trigger system.

The electromagnetic end-cap calorimeters consist of two wheels which are located at the two side of the barrel calorimeter, as shown in figure 4.10 and figure 4.8. The end-cap region covers a larger pseudorapidity range of $1.375 < \eta < 3.2$.

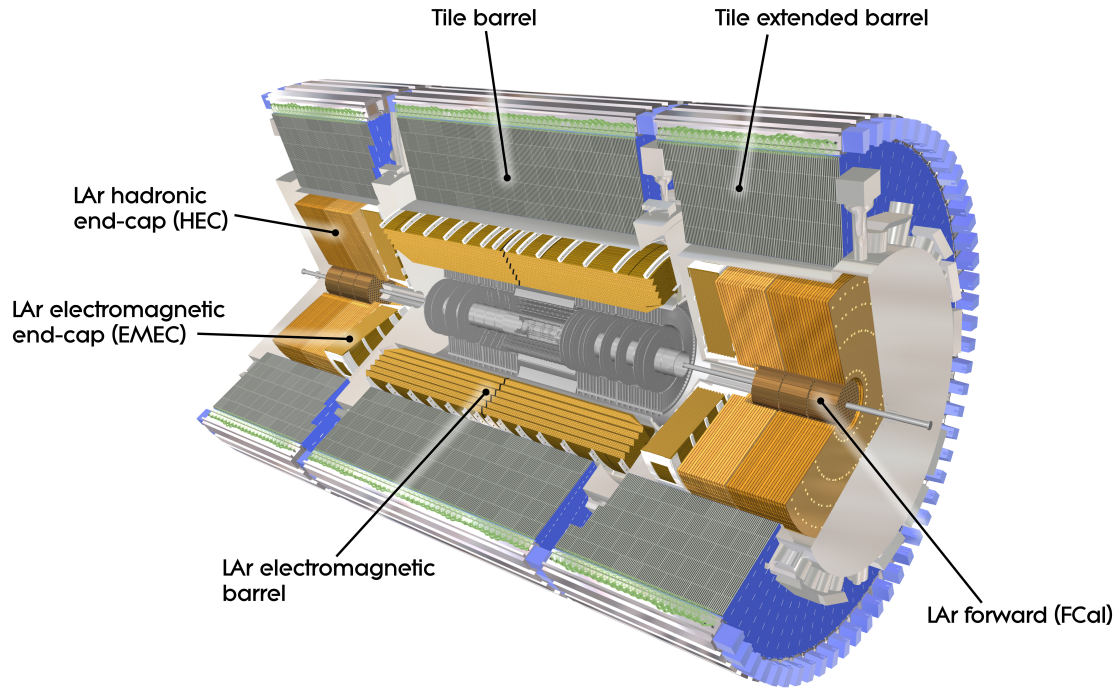


Figure 4.8: The cutaway layout of the ATLAS calorimeter [73].

4.2.4.2 Hadronic Calorimeter

The hadronic calorimeter consists two parts: the tile calorimeter and the LAr hadronic end-cap calorimeter, their locations are shown in figure 4.8 and figure 4.10. The hadronic calorimeter is designed to measure the energy of the hadrons that are not fully stopped by the LAr electromagnetic calorimeter and deposited energy in the hadronic calorimeter.

The tile calorimeter comprises layers of steel and acts as a sampling calorimeter. Showers are induced as particles hit the layers of steel. The tile calorimeter is also the scintillator, producing the photons and inducing electric current. The intensity of the current can be used to measure the energy of the hadronic particles. It consists of 420000 scintillating tiles that sample the energy deposited, it has a total weight of 2900 tonnes, which is the heaviest component in the ATLAS detector.

The hadronic calorimeter has lower precision than the electromagnetic calorimeter, it measures the energy and provides the location of the particles which interact via strong force. One part of the hadronic calorimeter is located in the barrel region and behind the LAr electromagnetic calorimeter. Another part consists of two extended barrels behind the end-cap calorimeters. The hadronic end-cap calorimeter is also using the liquid argon as the sampling material and cooled by the electromagnetic calorimeter's cryostat, and the copper and tungsten are chosen as the absorber materials.

4.2.4.3 Forward Calorimeter

The forward calorimeter detects particles that pass through the high pseudorapidity region, it covers the large pseudorapidity range of $3.2 < |\eta| < 4.9$. The schematic cutaway layout of the

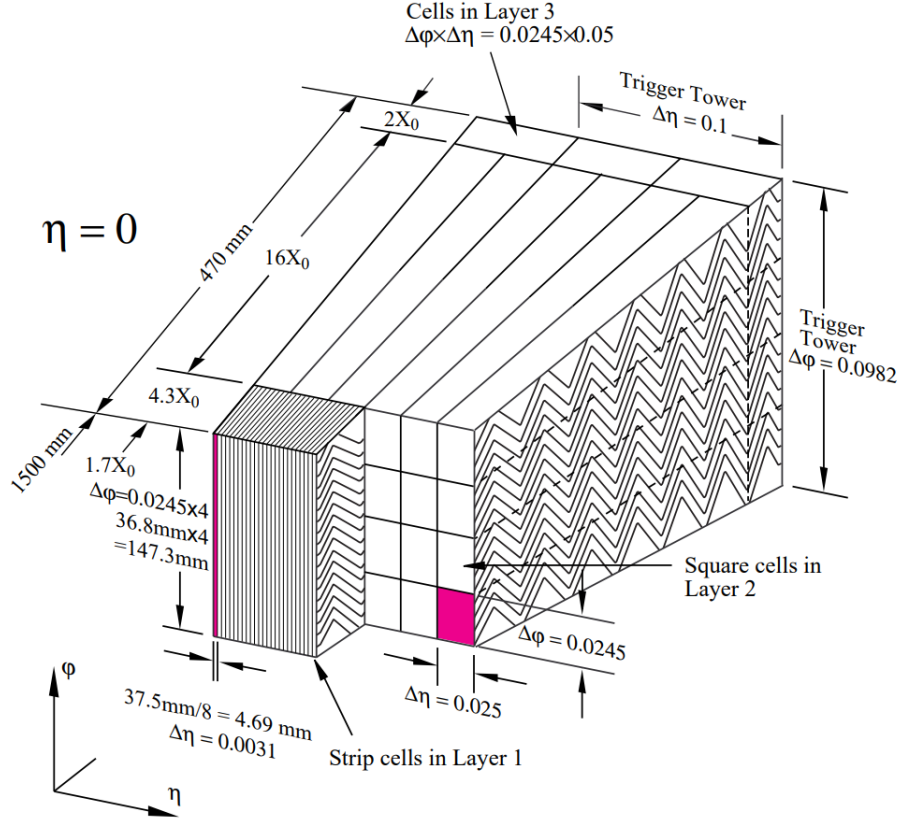


Figure 4.9: Geometry of the barrel electromagnetic calorimeter. For each of the three layers and the trigger towers, the granularity in η and ϕ of the cells is also illustrated [70].

forward calorimeter is shown in figure 4.10, and it is located far from the interaction point, resulting in less radiation damage. The forward calorimeter consists of three modules, with the module closest to the interaction point made of copper and known as the electromagnetic module, the second and third modules are made of tungsten.

Particles that can escape the detector easily are the neutrinos with high momentum and other weakly interacting particles, the large imbalance in the transverse momentum are reflecting their presence. The forward calorimeter are focusing on the physics with missing transverse energy and detecting the events with tagging jets due to its large coverage in the pseudorapidity and its location.

4.2.5 Muon Spectrometer

The muons from the proton-proton collisions can penetrate through the inner detector and calorimeters undetected, they will reach the outermost part of the ATLAS detector - the muon spectrometer, which is designed to identify the muons and provide the stand-alone high precision measurements of them. The muon spectrometer consists three parts: the toroidal magnets installed in both the barrel and end-cap region, the precision-tracking chambers and the trigger chambers, the cutaway view of the muon spectrometer is shown in figure 4.11.

The superconducting toroidal magnets can induce a strong magnetic field that allows the trajectories and the momenta of muons can be measured precisely. The muon spectrometer covers a pseudorapidity range of $|\eta| < 2.7$, the two end-cap toroidal magnets installed in both ends bend the tracks in $1.6 < |\eta| < 2.7$, while the magnetic field in the barrel region bend the tracks over the

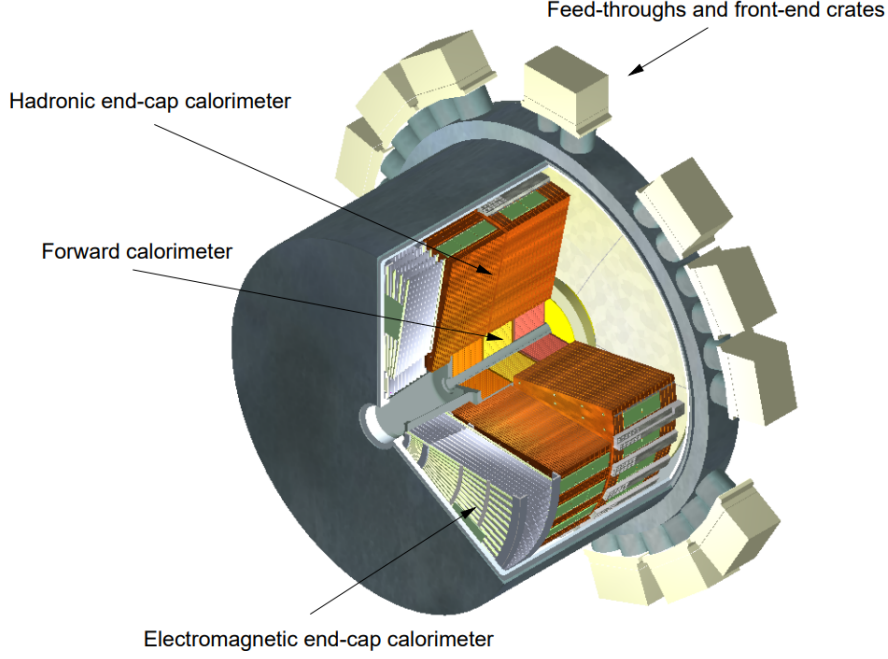


Figure 4.10: The cutaway layout of the forward calorimeter and end-cap calorimeters [70].

range $|\eta| < 1.4$. For transition region over the range $1.4 < |\eta| < 1.6$, the tracks are bent by both barrel and end-cap fields. In this way, the resolution degradation can be minimized by the magnetic fields normal to the trajectories.

4.2.5.1 Precision-tracking Chambers

The precision-tracking chambers are designed to provide high precision measurements on the curve of the tracks and location of the muons at the end of the detector. There are two types of precision-tracking chambers: the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). The CSC is a symmetric wheel located in the end-cap region, which is made of eight large and eight small chambers, the layout of CSC is shown in figure 4.12 (a). The MDT chambers are located in both the barrel and end-cap regions, their sizes are proportional to the distance from the interaction point. The shape of the chambers are rectangular in the barrel regions which is shown in figure 4.12 (c), while the chambers are wheels in the end-caps, all the chambers are made with three layers.

The CSC's are multi-wire proportional chambers with cathode planes segmented into strips in orthogonal direction as shown in figure 4.12 (b). The precision coordinate is provided by the cathodes with the strips normal to the wires, while the transverse coordinate is provided by the other cathodes with strips parallel to the wires, the tracks are reconstructed by the induced charges between the cathode strips. The CSC covers the range of $2 < |\eta| < 2.7$, and the innermost layer of MDT are replaced by the CSC's in this region. In the forward region, the chambers experienced the highest particle fluxes, therefore, the CSC is designed to have very small electron drift times and low neutron sensitivity.

In the MDT chambers, there are 354240 tubes filled with gas [75]. While these gas tubes interact with the muons which pass through, a trail of charged ions and electrons are leave behind that drift the sides and the centre of the gas tubes. The tracks of the muons can be reconstructed

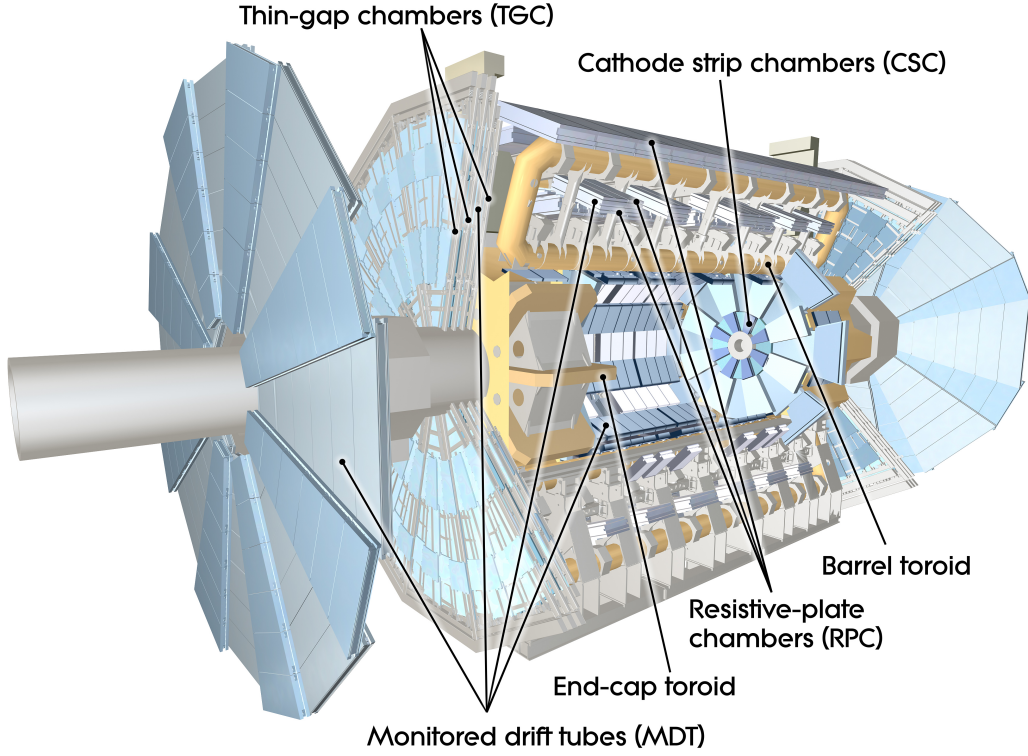


Figure 4.11: The cutaway layout of the muon spectrometer [74].

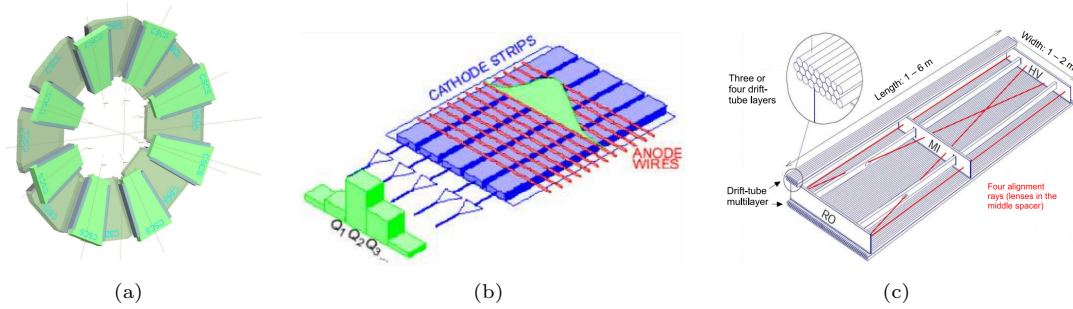


Figure 4.12: (a) The layout of the CSC with eight large and eight small chambers, (b) the segmentation of the CSC cathodes, (c) mechanical structure of a MDT barrel chamber [70].

by measuring the drift time of the charges, while the momentum of muons can be measured by the curvature. Figure 4.12 (c) shows the multilayer structure of the MDT barrel chamber, there are four tube layers in the innermost layer for enhancing the pattern-recognition function, while there are three tube layers in the other two multilayers.

4.2.5.2 Trigger Chambers

There are three levels of the trigger system in ATLAS, the first level (L1) trigger uses the data from the trigger chambers with reduced granularity, more details of the trigger system is discussed in section 4.2.8. There are two types of the trigger chambers: the Resistive Plate Chambers (RPC) in the barrel region and the Thin Gap Chambers (TGC) in the end-cap regions. The trigger chambers can provide high performance time resolution and the branch crossing in high precision, which is

one of the main requirements of the L1 trigger.

The trigger chambers cover a pseudorapidity range of $|\eta| < 2.4$ and a full azimuthal range. In the end-cap regions with high pseudorapidity, the trigger system requires a higher performance on the p_T -resolution and the granularity, due to the fact that the momentum is increasing with pseudorapidity η . In order to obtain the high p_T resolution, three layers of trigger detectors are used. Figure 4.13 shows the schematic layout of the muon trigger system, the low p_T triggers and the high p_T triggers use different layers of the trigger chambers, the RPC2 and TGC3 are taken as the reference planes. In both the barrel and end-cap region, the low p_T trigger information are provided by measuring the slope between the reference plane and the other plane and compare to the slope between the reference plane and the interaction point, while the high p_T triggers use the coincidence patterns from all the three layers.

The RPC is made of gaseous electrode-plate without wire, a 2 mm gas gap is formed in between two resistive plates which are separated by insulating spacers. A strong electric field is formed between two plates while high voltage is applied to them, charged particles are ionized and induced an electrical signal in the gas gap. Two layers of the PRC sandwich the MDT's in the barrel region as shown in figure 4.11, the two layers are known as the gas volumes which are read-out by two orthogonal strips.

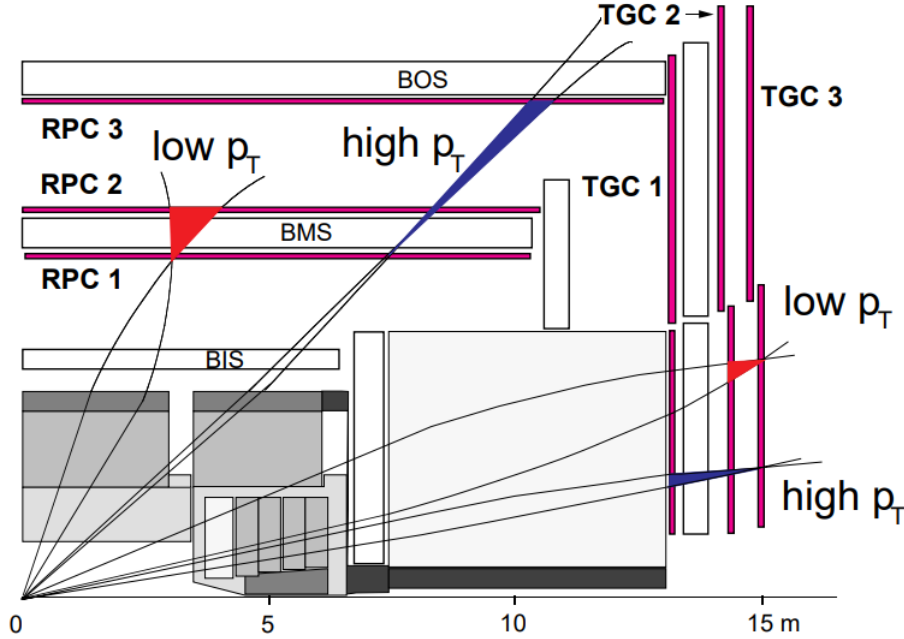


Figure 4.13: The schematic view of the muon trigger system [76].

The TGC is designed to provide the muon trigger capability and measure the second coordinate to complement the MDT's measurements. In the end-cap regions, the layers of the TGC are complemented the MDT wheel as shown in figure 4.11. The TGC's are multiwire proportional chambers, the wire groups measure the radial coordinate while the strips measure the azimuthal coordinate. The distance between two nearby wires is very small and a strong electric field is surrounding the wires, which provides a good performance on the time resolution and it is important for the end-cap detectors.

The tracks of the particles are not bent in the three trigger layers in the end-caps as the

layers are outside the magnetic field. In the transition region with $1.3 < |\eta| < 1.65$, the magnetic fields induced by the barrel and end-cap toroids are overlapped, inducing a strong inhomogeneous magnetic field in this region. On the other hand, low pseudorapidity region has very high particle fluxes with high momenta. These particles can pass through almost straightly, faking the tracks with high momentum. To reduce those backgrounds, the trigger in this region was excluded.

4.2.5.3 New Small Wheels

The increase in luminosity and energy in the Run-3 experiments challenge the background rejection performance of the ATLAS detector, some upgrades on the detector were planned to adopt the high luminosity. One of the main upgrades in the LS2 period is the installation of the New Small Wheels (NSW) in the muon spectrometer. The NSW incorporates two detector technologies: the resistive strip Micromegas detectors (MM) and the small Thin Gap Chambers (sTGC). The NSW is located in the forward region, specifically in the innermost part of the forward muon detectors. The schematic view of the location of MM and sTGC in the NSW is shown in figure 4.14. Each sector in the NSW consists of two sTGC wedges sandwich two MM wedges, with a spacer.

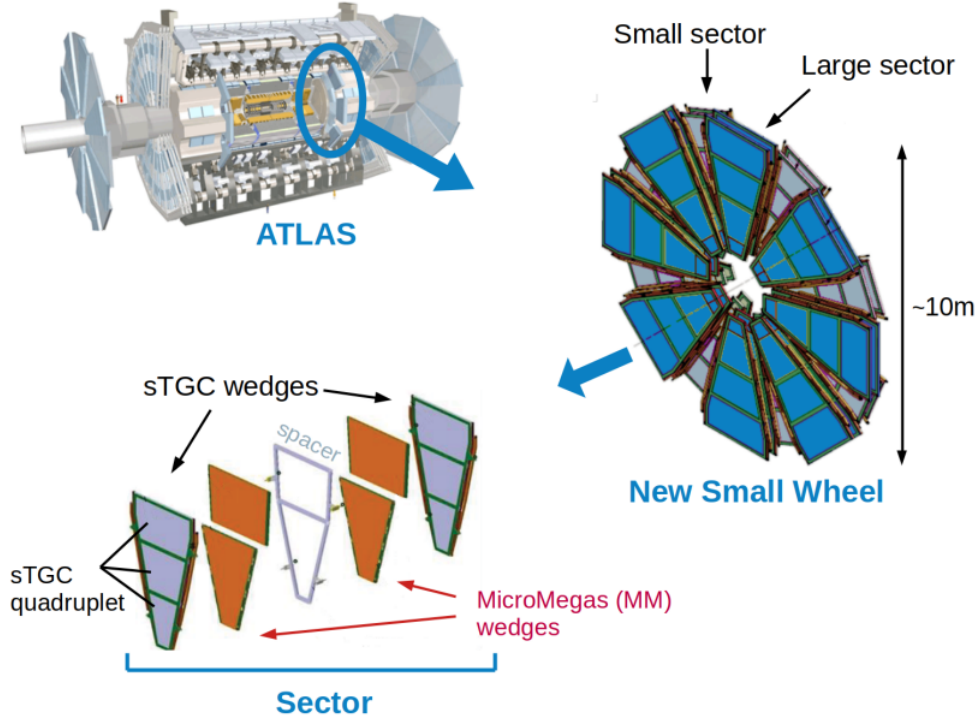


Figure 4.14: The schematic view of the location of NSW, and the location of the MM and sTGC in the NSW is also shown [77].

The NSW is designed to enhance the muon triggering and track reconstruction. The sTGC and MM are expected to provide better time resolution and angular resolution. The NSW has a symmetric structure with eight small and eight large sectors, its layout is shown in figure 4.14. These upgrades in the NSW enable a significant reduction in fake triggers and sufficiently increase the efficiency in handling the higher particle rate. The original small wheels do not have ability on the fake trigger rejection.

4.2.6 Magnet System

The magnet system in the ATLAS detector consists of four superconducting magnets, which has a cylindrical structure in overall with diameter of 22 m and a length of 26 m. The schematic view of the magnets is shown in figure 4.3, one solenoidal magnet is installed around the inner detector, and two toroidal magnets are installed in the both end-cap region while one toroidal magnet is located in the barrel region. The magnet system is designed to bend the charged particles by the Lorentz force $F = qvB$, the momentum of the particles can be determined by the curvature of the trajectories as $r = p/(qB)$.

The solenoidal magnet provides an axial magnetic field with strength of 2 T for the inner detector. This strong magnetic field ensures the trajectory of the high-momentum particles can be bent. As the calorimeter is installed outside the solenoidal magnet, the thickness of the magnet should be small enough for avoiding an impact on the performance of the calorimeter.

The barrel toroidal magnet has a symmetric structure, which consists of eight large coils and has a length of 25.3 m, which breaks a world record and being the largest toroidal magnet. The toroidal magnets can generate magnetic field up to 3.5 T for bending the trajectories of the muons.

The superconducting magnets are cooled using a flow of boiling helium. For keeping them in the superconducting state, the magnets should be cooled down to a temperature of 4.5 k. The barrel magnets have a cold mass of 370 tonnes, while each of the two end-cap magnets has a cold mass of 160 tonnes. In the case of quenching, the system is protected by using the enthalpy of the cold mass to absorb the stored energy of nearly 1 GJ.

4.2.7 Forward Detector

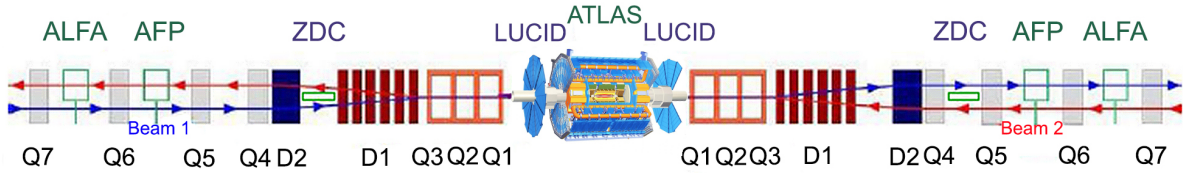


Figure 4.15: The schematic view of the ATLAS forward detector [78].

Four subdetectors are installed to complement the ATLAS detector in the very forward region to measure the particles at very small angles. Figure 4.15 shows the location of the four forward detectors in order relative to the ATLAS detector. The first system is the Luminosity measurement using Cerenkov Integrating Detector (LUCID), which is located at 17 m from the interaction point. The second system is the Zero Degree Calorimeter (ZDC), which is located at 140 m from the interaction point, right at the location where the two beams are split back into two beam pipes. The third system is the Atlas Forward Proton (AFP), which is designed to tag diffractive events and located at 210 m from the interaction point. The forth system is the Absolute Luminosity For ATLAS (ALFA), installed in a far away location of 240 m from the interaction point [78].

LUCID is specifically designed to measure the inelastic proton-proton scattering events in the very forward region and determine the luminosity with very small uncertainty, it is a Cerenkov detector dedicated to provide the online luminosity monitoring and measure the integrated luminosity. LUCID determine the luminosity via counting the number of interactions per beam-crossing, which is proportional to the number of detected particles. The current used ATLAS luminosity detector is the

LUCID-2 detector, which is still serving ATLAS during the Run-3 period. In the HL-LHC project, as the expected number of interactions per branch-cross is sufficiently increased, the upgraded detector - LUCID-3 will be used to meet the challenges.

The ZDC is primarily designed to measure the forward neutrons from the heavy-ion collisions, it has very large pseudorapidity coverage over the range $|\eta| > 8.3$. There are four modules installed in each arms in the ZDC: one electromagnetic module and three hadronic modules. The ZDC is also a sampling calorimeter, which detects the Cherenkov light induced by the high-momentum charged particles. The ZDC can provide separate sampling and readout for the measurements of energy and position, which reduces the shower sampling uncertainties for the measurements of the energy.

One of the main goals of the AFP detector is to identify the protons from the collisions with large pseudorapidity, it is located at 210 m from the interaction point and only 3 mm from the beam axis [79]. The AFP detector is installed with a pixel silicon tracker combined with a Time-Of-Flight system [79].

The ALFA detector is designed to measure the elastic proton scattering at a extremely small scattering angles of $3 \mu\text{rad}$. The absolute luminosity can be determined by amplitude of these elastic scatterings in the very forward region, and ALFA is, therefore, the farthest detector from the interaction point.

4.2.8 Trigger and Data Acquisition

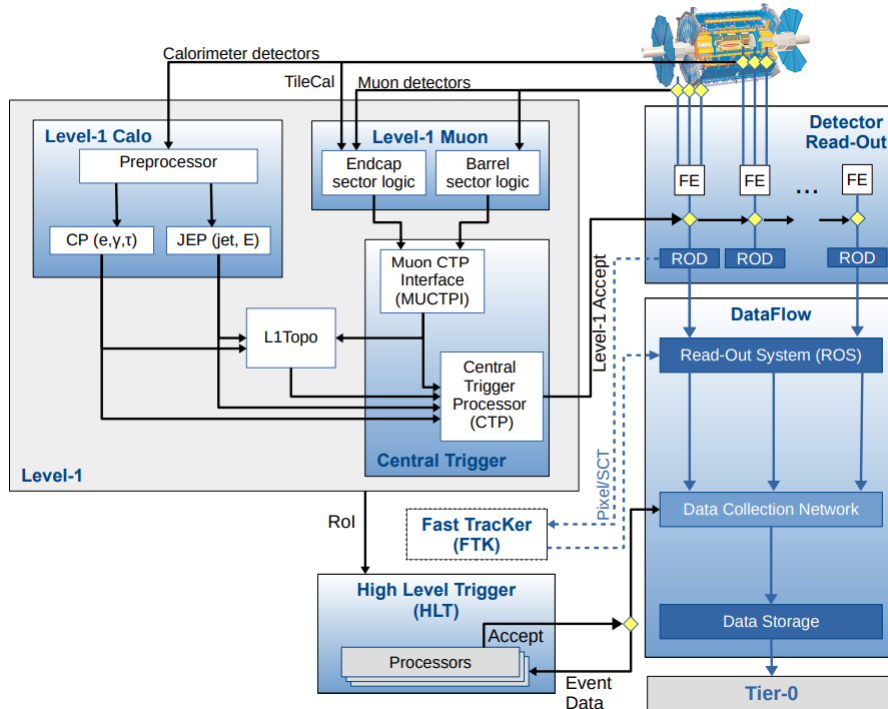


Figure 4.16: The ATLAS TDAQ system used in Run-2 experiments. [80].

The full Run-2 data with an energy of 13 TeV correspond to an integrated luminosity of 139 fb^{-1} , about 1.7 billions of proton-proton collisions happen per second in the ATLAS detector [81], which produced an extremely large data volume. The ATLAS trigger system is designed to filter the data events, and only about 1000 events were selected each second, which is an accessible level

of the data flow. The data acquisition (DAQ) system receives and buffers the event data from the readout electronics at the trigger system, the trigger system and the DAQ system incorporate to form the Trigger and Data acquisition (TDAQ) system in ATLAS. The TDAQ system selects interested events based on a trigger menu, deciding which events to be saved in the disk.

The ATLAS trigger system consists of three level: the level-1 (L1) trigger, level-2 (L2) trigger and the event filter. The L1 trigger is implemented as hardware based, using custom electronics. While the L2 trigger and event filter are software based and form the High-Level Trigger (HLT) system.

The L1 trigger is designed to searches for the high-momentum physical objects such as leptons, photons, and jets. As mentioned in section 4.2.5.2, the RPC and TGC provide the reduced-granularity information for the L1 trigger to search for the high-momentum muons, while the calorimeters are used for leptons, jets, events with large missing transverse energy and large transverse energy. The maximum acceptance rate of the L1 trigger is up to 100 kHz, the L1 trigger must have an extremely fast reaction on deciding whether to keep or reject an event.

The block diagram of the TDAQ system used in Run-2 experiments is shown in figure 4.16, the events passing the L1 trigger are sent to the HLT system. The HLT system is seeded by the Region of Interests (RoI) found in L1 trigger and perform the reconstruction of physical objects such as electrons, tracks, etc. The L2 trigger operates based on about 40000 CPU cores [81], reducing the event rate to accessible level for the offline analysis. The L1 trigger takes the data from the muon spectrometer and calorimeters, while the L2 trigger uses the data from the entire detector.

The data acquisition systems use point-to-point links with high speed for transferring the data and switching the networks. The readout system buffers front-end event data from the subdetectors electronic systems and made them available for sampling by the algorithms running in the HLT, and those HLT accepted events are sent to the CERN permanent storage via the data logger. The full logic is shown in figure 4.16.

In order to meet the challenges in the Run-3 experiments with a much higher luminosity, there are large upgrades on the ATLAS TDAQ system. The L1 trigger upgrades include the LAr energy super cells, which improves the trigger resolution and efficiency for searching the physical objects, and a major upgrades on the L1Calo system. The Fast TracKing (FTK) is implemented, allowing the fast hardware-based track finders to transfer full event tracking information to the HLT.

4.2.9 Summary of Second Upgrades

During the second long shutdown (LS2) period of the LHC, the ATLAS detector underwent several upgrades. Some of details are provided throughout this chapter, and this section serves as the summary of these upgrades.

The inner detector, described in section 4.2.3, is used for tracking the charged particles, and it is upgraded with a new system called the Inner Tracker (ITk). The ITk is expected to improve the tracking efficiency and position resolution.

The Liquid Argon (LAr) Calorimeter, described in section 4.2.4.1, is responsible for the energy measurement of the charged particles. The front-end electronics were replaced and upgraded for improving the readout capabilities.

The upgrades on the muon spectrometer is described in section 4.2.5.3, the MM and STGC were installed in the muon system for improving the triggering and tracking.

The Trigger and Data Acquisition system, described in section 4.2.8, is upgraded for improving

the front-end electronics such that the increase in the data rates can be handled. The readout system and the data processing is also upgraded for increasing the efficiency and capability.

Chapter 5

Experimental Data and Simulation Samples

The data obtained from the detector and passed through the trigger are in RAW data format. They are not yet ready for physics analysis. The data must be monitored and recorded during the data preparation phase. These processes are crucial as they provide information about data quality, luminosity, metadata, and the backgrounds that affect the final collision states.

The Monte Carlo (MC) technique is employed to simulate collisions and the detector. Simulated events are utilized for optimizing event selection, modeling the Higgs LFV signal, and some background processes. The MC samples are also utilized as inputs for multivariate classifiers and for setting the systematic uncertainties.

The flow of data and MC simulation is illustrated in Figure 5.1, which also indicates the format of the respective outputs. This chapter discusses ATLAS data processing and MC sample generation.

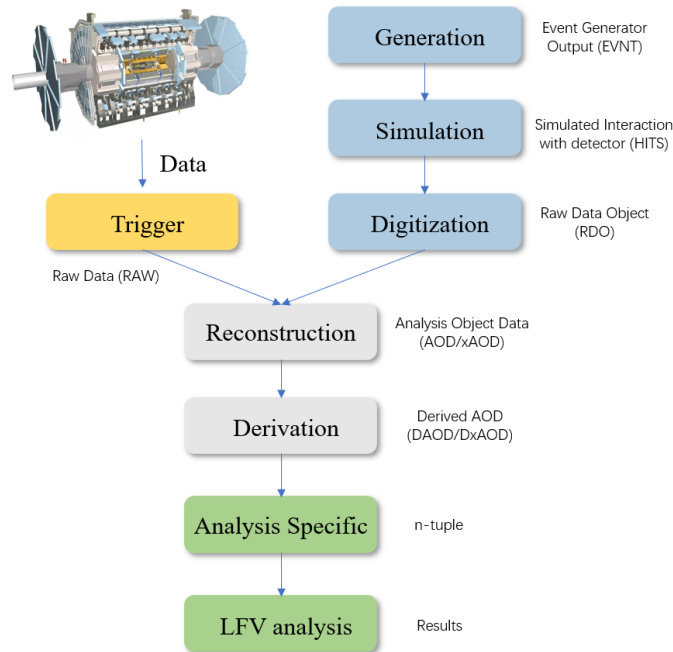


Figure 5.1: The processing flow and the data format in ATLAS.

5.1 Data Processing

Not all the data collected in the ATLAS detector is used in the physics analysis. Data quality must be ensured through a monitoring procedure and expert reviews. The nominal Run-2 operations workflow for data quality assessment is shown in Figure 5.2. The data that pass the trigger will be input into the Tier-0 stage, which involves prompt reconstruction. The data from the Tier-0 computing facility will be sent to different processing streams and reviewed by experts. Meanwhile, the online histogram sources, which include events accepted by the HLT, will be sent to the main control room and monitored by shifters who must complete the shift training session. The data is iterated in the Conditions Database loop and reprocessed in Tier-1 with improved software. This monitoring algorithm ensures that the reconstructed physical objects are acceptable and that data quality defects are rejected.

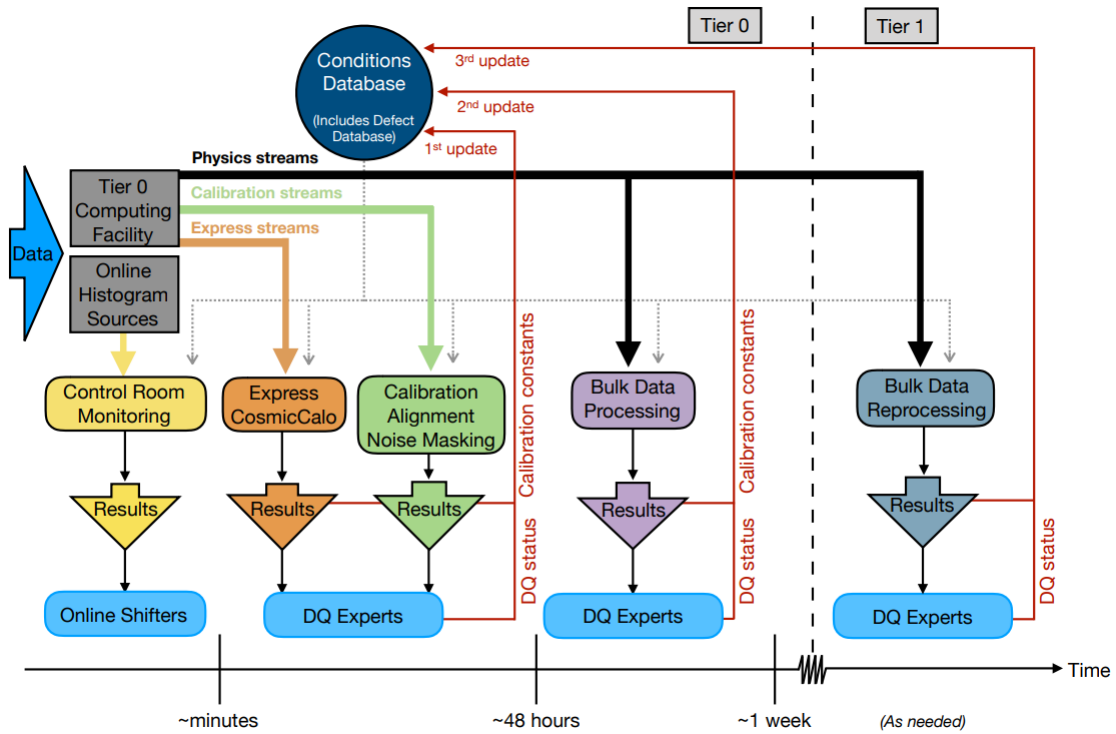


Figure 5.2: The ATLAS run-2 workflow for the data quality assessment system [82].

In order to increase the possibility of a collision, the beam size has to be squeezed as much as possible. There are hundreds of billions of protons in a bunch, with a size nearly equivalent to the width of a human hair. However, only tens of collisions occur per bunch crossing. The time interval between each bunch crossing in one continuous data recording period is 25 ns [82], resulting in approximately several hundred million collisions per second. The beams usually circulate in the LHC for hours within a single data recording period, also known as a fill.

In LHC Run-2 operation, the number of interactions per crossing is recorded as integrated luminosity. An individual run, typically corresponding to a fill, is further divided into luminosity blocks. Within one luminosity block, the data are recorded under similar quality conditions within a minute. All luminosity blocks within this 60-second period are summed up and defined as a fill, while a group of fills taken under very similar conditions is defined as a single period of operation. A full LHC operation, such as Run-2, is defined as the sum of these periods. During the data-taking

periods, the luminosity is recorded.

Not all the data delivered from the LHC are recorded by the ATLAS detector. In the data-taking period, when stable collision conditions have been achieved, the ATLAS detector does not record data, as its subsystems are still "warming up" by high voltage [82]. ATLAS detector started to record data after the declaration of "ready for physics". For those data used in the physics analysis, they are "good for physics" and the integrated luminosity are calculated from the luminosity blocks which are selected based on the *Good Runs Lists* (GRL), events in the GRL are those candidates pass the data assessment system. Figure 5.3 (a) shows the total integrated luminosity in each year and month. For the data-taking campaign of each year during the Run-2 operation period and the proton-proton collisions with centre-of-mass energy of 13 TeV, the recorded integrated luminosities in the four years operation from 2015-2018 are 3.63 fb^{-1} , 35.5 fb^{-1} , 46.4 fb^{-1} and 60.0 fb^{-1} [82], respectively. The total integrated luminosity used in the analysis is 138.42 fb^{-1} . Figure 5.3 (b) shows the data quality performance as a function of integrated luminosity for each year, a total efficiency of 95.6% is obtained for the full Run-2 dataset with $\sqrt{s} = 13 \text{ TeV}$ [82]. It is worth mentioning that the muon subsystem achieved a high data quality performance with efficiency around 99.8%, and the upgrades in LS2 is expected to further improve its performance in the Run-3 operation.

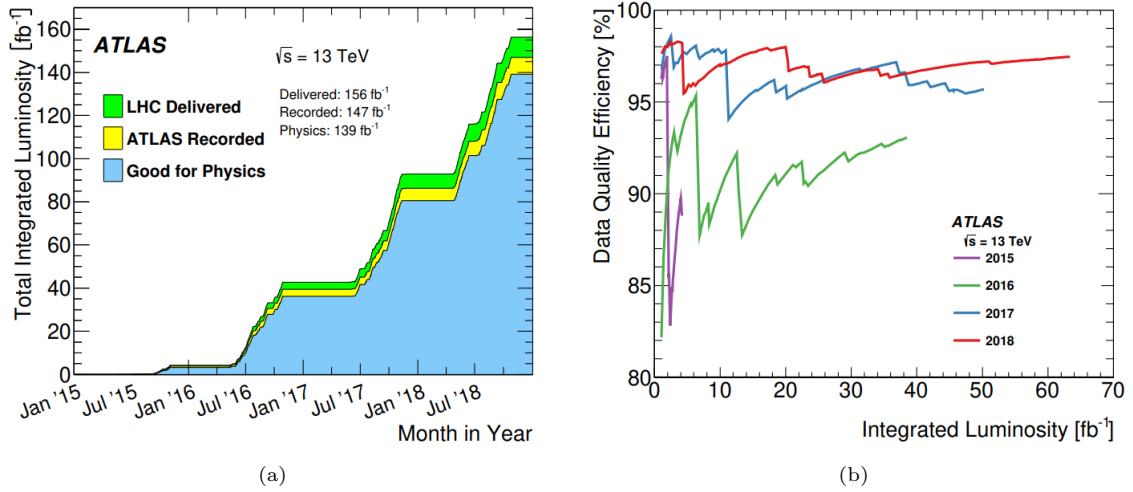


Figure 5.3: (a) The integrated luminosity recorded by ATLAS detector in the Run-2 operation period, the delivered, recorded and "good for physics" fraction are illustrated. (b) The changes in the cumulative data quality efficiency with the integrated luminosity for each data-taking campaign [82].

This analysis uses the LHC full proton-proton Run-2 dataset, with an integrated luminosity of 138.42 fb^{-1} after passing the data quality assessment system, and the corresponding time interval between bunch crossing is 25 ns. Events are further selected based on the trigger selections which is discussed in section 8.

5.2 Monte Carlo Simulation Samples

The MC-template method is based on data-corrected simulation samples. In the MC-template method, MC simulated events are used to model most of the backgrounds from Standard Model (SM) processes, as well as the signal processes of $H \rightarrow \ell\tau$. In the symmetry method, the analysis exploits

a data-driven technique. The MC samples are used to model the signal processes, optimize event selections, calculate the electron detection efficiencies, fine-tune the fake background estimation, and model the systematic uncertainties.

All MC samples were produced with the ATLAS simulation infrastructure [83] using the full detector simulation performed by the GEANT4 [84] package. The simulated events are reconstructed with the pile-up events, which refer to multiple interactions in the same and nearby bunch crossing by overlaying the minimum-bias events.

The full MC simulation process covers the event generation, simulation, and digitization, as shown in figure 5.1. The event generation process runs with the software of the *Athena* framework [85], which is a powerful package involving the offline software system of ATLAS experiments. The background and the LFV signal events are generated via this framework. However, the generators are written with external packages and employed by the authors for their maintenance. The simulation packages are summarized in table 5.1, including the event generators used for each process, the parton distribution function (PDF) sets for the hard process, the set of tuned parameters, and the order of cross-sections used in normalizing the events. The details are discussed in the following sections, specifically for different production modes and background processes.

5.2.1 LFV Signal Samples

The MC simulated events are used to model the signal processes of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$, based on the Higgs boson production modes as introduced in section 2.4. The considered production modes used in the MC-templated method are the gluon-gluon fusion (ggF), vector-boson fusion (VBF) and the vector boson associated (VH) production, and the top-quark pair ($t\bar{t}H$ production has negligible contribution). The Higgs boson mass used in the calculation of the expected cross-sections and branching ratios is the measured value of 125 GeV, and the cross-sections are fixed to the SM predictions [86] throughout the analysis.

The details of the event generations in each production modes are introduced below:

- **Gluon-gluon fusion (ggF)**

The ggF production mode is simulated at next-to-next-to-leading-order (NNLO) accuracy, using the POWHEG NNLOPS [87, 88] program. The POWHEG is used for large-scale production of top, W and Z events, it used hard-scattering processes which estimated under the next-to-leading (NLO) order QCD perturbation theory. For the matrix elements (ME), the generator is POWHEG BOX v2. The simulations of parton showers (PS) run with the PYTHIA 8 [89] program, which is the default generator in ATLAS [83], and it can be used to fix the scale at which partons are given to hadronization.

The estimation of the cross sections at the LHC relies on the knowledge of the momentum distribution functions of the partons within a proton, which is described by the parton distribution functions. The PDF is defined in terms of the matrix elements of certain operators. The simulations of ME and PS in the PDF run with the PDF4LHC15NNLO [90] and CTEQ6L1 [91], respectively. The generators have to be tuned in order to adopt the LHC environment, the AZNLO [92] parameter is used for tuning.

The normalization of the gluon fusion prediction in the MC events is at the next-to-next-to-next-to-leading order (N³LO) cross-section accuracy in QCD process plus electroweak corrections at next-to-leading order.

- **Vector boson fusion (VBF) and associated production (VH)**

The Higgs boson production mode via the vector boson fusion is generated also with the POWHEG BOX v2 at the NLO accuracy, interfacing with PYTHIA 8 for parton showers and tuned also with the AZNLO parameters set. The PDF set used for the ME is the PDF4LHC15NNLO, and similar to the ggF, CTEQ6L1 is used for the PS.

The same generators and simulations are also applied to the VH production mode. In both production mode, the MC prediction is normalized to cross sections at next-to-next-to-leading order (NNLO) in the QCD process plus the electroweak corrections at NLO.

- **Top fusion ($t\bar{t}H$)**

This production mode is not considered in either the MC-templated method or symmetry method, but this process could be the potential signal in the Run-3 experiments and beyond. The Higgs boson production via a pair of top-quark is simulated using the POWHEG BOX v2 generator at the NLO accuracy with the NNPDF3.0NNLO [93] PDF set. The simulated events are also interfaced to PYTHIA 8 but using the A14 [94] tune and the NNPDF2.3LO [95] PDF set.

Process	Generator		PDF		Tune	Cross Section Order
	ME	PS	ME	PS		
Higgs boson production modes						
ggF	POWHEG BOX v2	PYTHIA 8	PDF4LHC15NNLO	CTEQ6L1	AZNLO	N ³ LO QCD+NLO EW
VBF	POWHEG BOX v2	PYTHIA 8	PDF4LHC15NNLO	CTEQ6L1	AZNLO	NNLO QCD+NLO EW
VH	POWHEG BOX v2	PYTHIA 8	PDF4LHC15NNLO	CTEQ6L1	AZNLO	NNLO QCD+NLO EW
$t\bar{t}H$	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO QCD+NLO EW
Background processes						
V+jets	SHERPA 2.2.1 [96]		NNPDF3.0NNLO		SHERPA [97]	NNLO QCD+LO EW
V+jets*	POWHEG BOX v2	PYTHIA 8	CT10NLO [98]	CTEQ6L1	AZNLO	NNLO
Diboson	SHERPA 2.2.1		NNPDF3.0NNLO		SHERPA	NLO
$t\bar{t}$	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO	NNPDF2.3LO	A14	NNLO+NNLL
Single top	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO

Table 5.1: Summary of the MC generators used to describe the signal and background processes, the V+jets with * symbol denotes additional samples used by the symmetry method.

The simulation of the LFV decays of the Higgs boson: $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ are modelled using EVTGEN v2.0 [99], and the tau lepton decays are modelled by the generator of TAUOLA [100].

5.2.2 Background Samples

The MC simulated events are used to model the background processes, which involves the SM processes of $H \rightarrow \tau\tau$ and $H \rightarrow WW$. They are simulated using the same generators and programs as those used in the LFV signal samples, but including the production mode of $t\bar{t}H$ as the background process. Other background processes involve electroweak production of W/Z bosons via vector boson fusion, the Drell-Yan production of W/Z +jets, the diboson process, the top-quark pair production and the single top-quark production. The details are summarised in figure 5.1 and listed below:

- **W/Z boson in association with jets**

The vector bosons may be produced in the heavy-quark decays, quark-quark interactions or quark-gluon interactions from the hadron collider. The production of the vector bosons usually in association with one or more jets. The Z boson could further decay into a pair of τ lepton or a pair of light leptons (ℓ), those decays of $Z \rightarrow \tau\tau$ +jets and $Z \rightarrow \ell\ell$ +jets are the main background processes in this analysis.

The production of the V +jets process is simulated with the SHERPA v2.2.1 generator to NLO matrix element for up to two partons, and leading order matrix elements for up to four partons. In symmetry method phase space, the simulation with the generators of POWHEG BOX v2+PYTHIA 8 can provide more statistics: these samples were produced inclusive at NLO in perturbative QCD with additional parton emissions considered only at the parton shower level, which results in most of the statistics being located at low p_T of the vector bosons, and this is the case in the symmetry method phase space.

The simulated events are interfaced to SHERPA using the SHERPA tune and the NNPDF3.0NNLO PDF set, the samples are normalised to a NNLO prediction [101]. Electroweak production of the final states with two jets are generated using the same program of SHERPA v2.2.1, using leading order matrix elements with up to two additional parton emissions.

- **Diboson**

In the diboson process, a pair of vector bosons are produced from the quark-quark interactions, through exchange of a virtual quark or another vector boson. The productions of diboson process are simulated with SHERPA v 2.2.1 generator to NLO matrix elements for up to one additional parton and at leading order for up to three additional parton emissions. The simulated events are interfaced to SHERPA using the SHERPA tune and the NNPDF3.0NNLO PDF set, the samples are normalised to a NLO prediction.

- $t\bar{t}$

Top quarks are usually produced in pairs as $t\bar{t}$ process, and top-quark can decay into a bottom-quark and a W boson, therefore, the quark pairs can contribute to the diboson background process.

The productions of the $t\bar{t}$ process are simulated with the POWHEG BOX v2 generator at NLO accuracy with the NNPDF3.0NNLO PDF set. The events are interfaced to PYTHIA 8 using the A14 tune, the samples are normalised to a NNLO+NNLL prediction.

- **Single Top**

The single-top production is modelled using the POWHEG BOX v2 generator at NLO in QCD with the NNPDF3.0NNLO PDF set. The events are interfaced with the PYTHIA 8 using the A14 tune, the samples are normalised to a NLO prediction.

5.2.3 Detector Simulation

Simulation of the ATLAS detector provides information of the detector response for different physics processes. The simulation process uses the outputs from the event generators to mimic the response of the true detector. The outputs of the simulation process are saved in the format of *HITS*, which contains the information of position, time and the energy deposited in the active region of the detector.

All the MC samples are processed through the full ATLAS detector simulation based on the GEANT4, which is a toolkit for the simulation of the passage of particles through matter. The full simulation chain is integrated into the Athena framework, which is an offline software framework for the ATLAS experiments. The geometry description of the ATLAS detector is also included in this framework and used in the full simulation chain. Using the simulation infrastructure [83], the run conditions and geometry databases can be monitored.

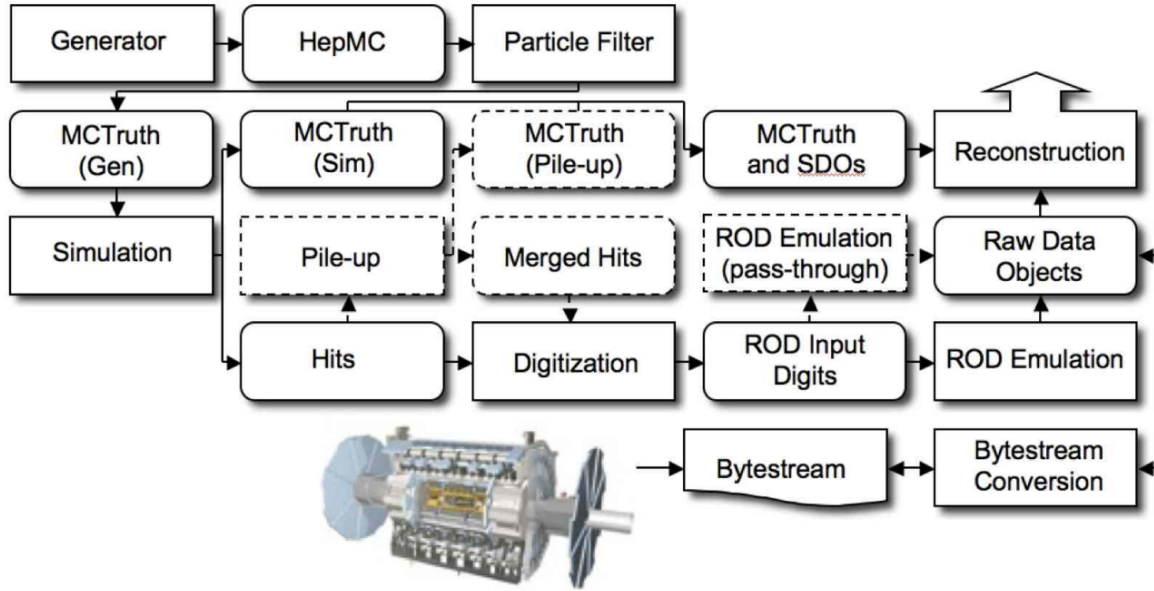


Figure 5.4: The ATLAS simulation software flow, from event generation to the step of reconstruction. The optional pile-up portion of the chain is dashed [83].

The full simulation algorithm used in ATLAS is shown in figure 5.4, more details of the flow is shown in figure 5.1. Generators are used to produce events in *HepMC* format, these events will be filtered with certain property. The information of incoming and outgoing particles in an event is recorded as *truth* in both the event generation and detector simulation. Each particle is recorded even if it is not passed through the detector. The truth include the tracks information of the particles passed through the ATLAS detector, and those information are stored in the outputs of the simulation, which is in the format of Simulated Data Objects (SDO). The truth is also the way to connect hits to tracks, the McTruth from the generator is the only source which is used by the detector simulation.

5.2.4 Digitization

The digitization converts events into a format in response to the detector; the outputs are saved as *digits*. The energy deposited is digitized into voltages and currents for comparison with the readout of the ATLAS detector. The digitization process takes the hits from the simulated events as input and converts them into digits, which are then input to the Read Out Drivers (ROD) in the detector electronics. The Raw Data Objects (RDO) that are input to the reconstruction step are produced by the ROD emulation. The outputs from the digitization are in the same format as the output of the ATLAS TDAQ system. The data from the detector in bytestream format is converted into the RDO format and then processed in the reconstruction packages along with the simulated

events that follow. The details of the object reconstruction are discussed in chapter 6.

The dashed components in figure 5.4 correspond to the simulation of pile-up. To simulate pile-up, hits from each type of event are overlaid. They can be configured at runtime, typically including signal, minimum bias, cavern background, beam gas, and beam halo events [83]. The number of events to be overlaid per bunch crossing is controlled and recorded by a pile-up manager service.

The data are reconstructed into Analysis Object Data (AOD) in the reconstruction step, usually resulting in a huge size about several petabytes. The derivation step discards the information which is not used in the objective analyses and converted the data into Derived AOD (DAOD) format. In this analysis, the main input type of data format is the *DAOD.HIGG4D1*, which is the dataset input to the software package that can be used to produce *n-tuple*, which refers to the data saved in the root file that can be used in the physics analysis. The software package of xTauFramework is used for the n-tuple production in this analysis.

5.3 Muon Software Validation

The ATLAS detector is a complex machine, its performance need to be understood and validated by the full detector simulation, which is developed by the teams responsible to each of the subdetector [83], including the inner detector, calorimeters, muon spectrometer and magnet system. In this section, the implementation of the Run-3 validation algorithm and performance monitoring procedure for the muon spectrometer is introduced.

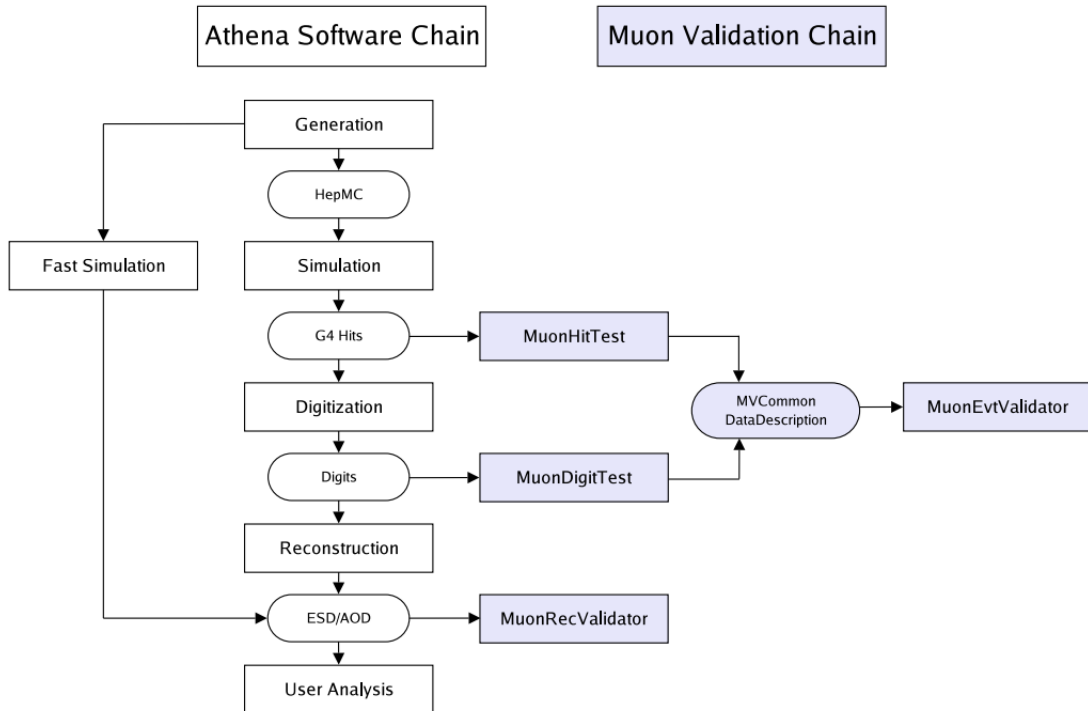


Figure 5.5: The flow of the muon validation chain, which is implemented into the Athena software chain. G4 Hits refer to GEANT4 Hits, ESD refer to Event Summary Data and MV refer to Muon Validation [102].

The simulation of muon spectrometer and the muon validation algorithm is integrated into the ATLAS offline software framework known as Athena, which is primarily developed by the LHCb

experiment and based on the GAUDI architecture. Figure 5.5 shows the workflow of the system combines the Athena software chain and muon validation chain. The muon validation algorithm is implemented in the simulation, digitization and reconstruction processes for monitoring the performance and plausibility of the respective transformation job outputs. As mentioned previously, the output from the GEANT4 simulation of each interaction point of the muon passing through the active component of the detector is known as a *Hit*. The *digits* model the response of the detector, while the AOD outputs include the information of the reconstructed muon track from digit-collection.

The geometry of the muon spectrometer is described by the MUONGEOMODEL package in Athena framework, while the muon detector description provides the information of each station and readout element of the muon subdetectors. The validation algorithm can output the variables that are defined by retrieving the station and readout geometry information, the performance of muon subdetectors are then monitored by checking the histograms of those validation variables in a bin-by-bin approach. The Run-3 validation algorithms also apply to the new small wheels, there are six validation algorithms for each of the subdetector: CSC, TGC, MDT, RPC, MM and sTGC.

5.3.1 Muon Simulation Hits

Muon hits are generated when the charged particles pass through a sensitive part of the muon subdetectors, which are described by the GEANT4 simulation. The GEANT4 simulation of the interaction between the tracks and the sensitive part of the muon chambers is encapsulated as a *hit object* or GEANT4 hit. In other word, the hits represent the energy deposit by a particle in the active part of the detector. The geometry information of each hit is stored by labelling a *Hits Identifier* (HitId) to each of them. The hit objects are decoupled from the GEANT4 geometry, but the geometrical information can be accessed via the identifiers. The identifiers are 64-bits (or 32-bits) integers, which cover the fields that define the readout elements of each of the subdetector.

The HitIds of the hits can be retrieved via the *HitIdHelper* classes implemented in the MuonSimEvent package in the Athena framework. The common station information of the hits such as the name ¹ (stationName), the η sector (stationEta) and the ϕ sector (stationPhi) are the main variables for monitoring the performance of the muon subdetectors. The station information can be retrieved by the HitIds via the HitIdHelpers. Variables corresponding to the muon technology can also be retrieved from the identifiers. For example, the gas gap information of TGC, the chamber layer and wire layer information of CSC. The hits carry not much information as the geometrical information is encapsulated in the HitIds, but kinematic information such as step length, kinetic energy, energy deposit and the particle encoding (pdgId) of a particle can be retrieved from the hits.

In addition to the HitIds, the muon subdetector offline identifiers are used for labeling the muon digits and accessing the information from the readout geometry provided by the Muon Detector Description (MuonDetDescr) package. The specific offline identifiers can be retrieved via the IdHelpers in the MuonIdHelpers package, with corresponding variables such as station and layer information (obtained from the HitIds). The readout elements contain all the geometrical information related to clients, such as the number of tubes (layers, wires, strips), variables used to define the active surface of the subdetectors, or variables used to check the detector acceptance. The readout elements of each subdetector also provide the position information from the global

¹The stationName refers to an unsigned integer maps to different sectors. For example, the CSC Short ϕ sector (CSS) and CSC long ϕ sector (CSL), etc.

and local reference systems of the subdetector. The global reference system refers to the standard ATLAS right-handed coordinate system centered at the interaction point, as shown in Figure 4.4. The latter refers to the local coordinate system, which is centered on each of the readout elements, with the x-z plane being the surface of the elements. The MuonReadoutGeometry package also provides the global-to-local (or local-to-global) transformation function for retrieving the position information from one system to the other. In the simulation process, this hit information can be retrieved by using the offline identifiers to access the readout geometry.

The GEANT4 hits are collected in six containers for each of the muon subdetectors using the *AthenaHitsVector*, and information from the identifiers is also exploited. The validation variables are defined by iterating through all the hits within the containers in a single loop. In this loop, the Athena message services are called to maintain and monitor the run status.

The outputs (histograms of the validation variables) from the loop are saved in n-tuple format and used for monitoring purposes, with the monitoring procedure introduced in Section 5.3.5. The variables used for the monitoring procedure are selected from among the candidates saved in the output n-tuple file. The selected variables for the CSC and TGC simulation validations are summarized in Table 5.2. The CSC and TGC Run-3 layouts in terms of the global x-y planes (plane view of the global position X and Y) for the simulation hits are shown in Figure 5.6, which were generated with 4000 $Z \rightarrow \mu\mu$ events to achieve higher resolution.

Identifiers	CSC	TGC
HitId	station η	
	station ϕ	
	chamber layer	gas gap
	wire layer	
Hit	kinetic energy	
	energy deposits	
	global time	
	local position X	
	local position Y	
	local position Z	
offline Id	global position X	
	global position Y	
	global position Z	
	strip	channel

Table 5.2: The list of the validation variables from the simulation hits output, which are monitored in the DCube-histogram monitoring system for the CSC and TGC subdetectors. There are two classes of variables obtained separately by using the HitId via the HitIdHelper and the offline Id via the IdHelpers. The variables below the dash line are obtained from the hits directly, including the HitId.

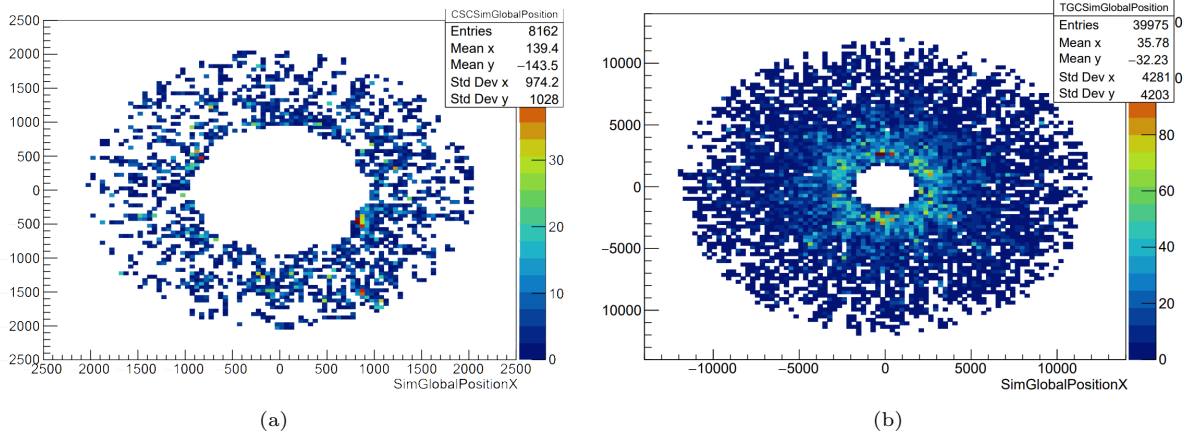


Figure 5.6: The x-y plane view of the (a) CSC and (b) TGC for the simulation hits, both were generated with the 4000 $Z \rightarrow \mu\mu$ events which were accessed on 12th October, 2021.

5.3.2 CSC and TGC Digitization

In the digitization process, hits are transformed into readout signals by modeling the detector response. The digitization process is introduced in Section 5.2.4. The muon validation of the digitization process consists of six algorithms, one for each muon subdetector. This section specifically focuses on the validation algorithms for the CSC and TGC.

The validation of the digitization process relies on the readout geometry and the offline identifiers. The HitId is associated with the hits; similarly, each digit is labeled by an offline identifier which contains the digital geometry information. Therefore, the digits have light contents (such as the global time), as the geometrical information is encapsulated in the identifiers. The resolution of the readout geometry is simulated in the digitization process.

As shown in Figure 5.4, the digits are converted into the RDO format and input into the reconstruction. The digits-to-RDOs conversion provides the simulation of the electronic output as raw data. In principle, any simulation information is lost after the digitization; both the digits and RDOs do not have a connection to the originally simulated particles. The HitId field is orthogonal to the offline identifiers associated with the digits, as the simulated event is disentangled from the digitization and reconstruction. The MuonSimData package can provide the collection of the simulated raw data that maintains the connection to the simulation for obtaining information about the MC truth tracks (MCTruth). The digitization algorithms can also produce Simulated Data Objects (SDOs), which contain the information about the contribution from the particle or noise to a hit and the amount of energy deposited. Each RDO is associated with an SDO in different ways that depend on the muon subdetectors.

The TGC has a wheel-like geometry, as depicted in Figure 4.11. The hits in the TGC are recorded with position information in the radial direction (r-direction) and azimuthal direction (ϕ -direction). The r-direction hit positions are measured using wire gangs in each element, while the hit positions in the ϕ -direction are determined using strips. In the TGC digitization process, the simulation takes into account various aspects of the detector response, such as the time response, signal propagation time along the wires and strips, and the detection efficiency of the wires and strips [103].

In the CSC digitization process, the incident charged particles that pass through the gases can

CSC Digitization Validation			
Variables	Digit	SDO	RDO
station η	✓	✓	✓
station ϕ	✓	✓	✓
global position X	✓	✓	✓
global position Y	✓	✓	✓
global position Z	✓	✓	✓
local position X	✓	✓	✓
local position Y	-	✓	✓
chamber layer	-	✓	✓
wire layer	✓	✓	✓
strip	-	✓	✓
barcode	-	✓	-
time	-	-	✓
measure ϕ	-	-	✓

Table 5.3: The list of the CSC validation variables from the digitization outputs, including the digit, SDO and RDO. The selected variables are monitored in the DCube-histogram monitoring system, which are obtained via the offline identifiers associated with the digit, SDO and RDO, respectively.

TGC Digitization Validation			
Variables	Digit	SDO	RDO
station η	✓	✓	✓
station ϕ	✓	✓	✓
global position X	✓	✓	✓
global position Y	✓	✓	✓
global position Z	✓	✓	✓
local position X	✓	✓	✓
local position Y	✓	✓	-
gas gap	✓	✓	✓
channel	✓	✓	✓
word	-	✓	-
barcode	-	✓	-
global time	-	✓	-

Table 5.4: The list of the TGC validation variables from the digitization outputs, including the digit, SDO and RDO. The selected variables are monitored in the DCube-histogram monitoring system, which are obtained via the offline identifiers associated with the digit, SDO and RDO, respectively.

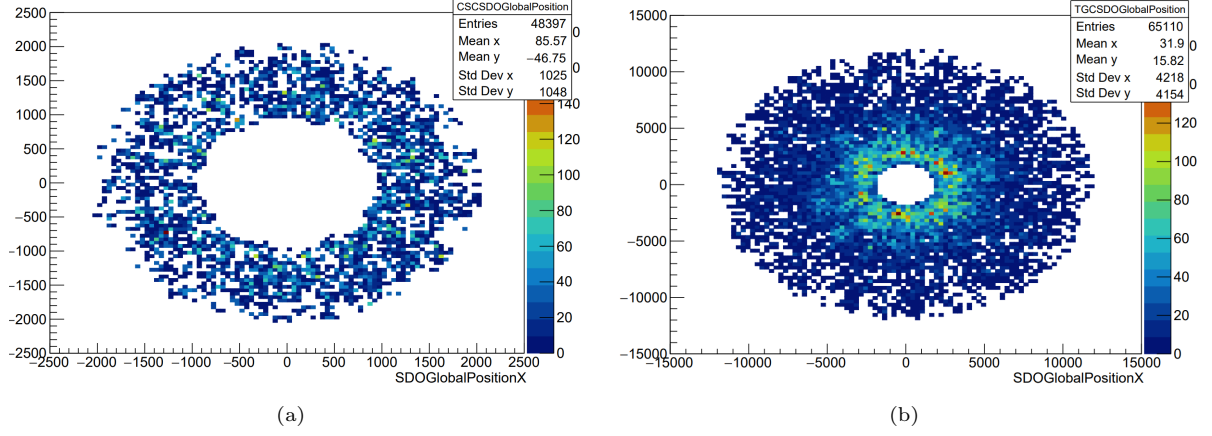


Figure 5.7: The x-y plane view of the (a) CSC and (b) TGC for the SDO outputs, both were generated with the 4000 $Z \rightarrow \mu\mu$ events which were accessed on 12th October, 2021.

induce ionization, the detector response of the avalanche on the anode wires caused by the induced electrons is simulated.

The digitization algorithms for each of the muon subdetector produced three outputs in the n-tuple format: digit, SDO and RDO. The selected variables for the digitization validation are summarised in table 5.3 for the CSC and table 5.4 for the TGC. The measures ϕ is 0 if r-direction is measured and is 1 if ϕ -direction is measured. The CSC and TGC Run-3 layout in term of the global x-y plane view for the SDO, digit and RDO outputs are shown in figure 5.7, figure 5.8 and figure 5.9, respectively. All the plots show a perfect shape with the desired symmetry.

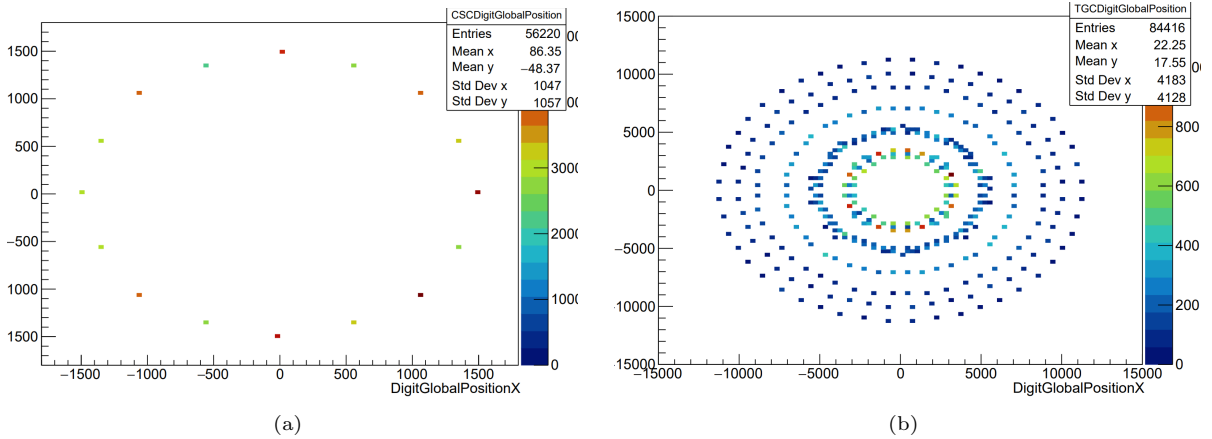


Figure 5.8: The x-y plane view of the (a) CSC and (b) TGC for the digit outputs, both were generated with the 4000 $Z \rightarrow \mu\mu$ events which were accessed on 12th October, 2021.

5.3.3 Muon Reconstruction

The raw data objects (RDOs) are not directly suitable for input to the reconstruction algorithms. Therefore, they need to be transformed into the Prepared Raw Data (PRD) class. The PRDs are the output of the muon validation chain and serve as input to the reconstruction process. The data in this class are formatted in a way that is ready for reconstruction. For example, drift time is converted

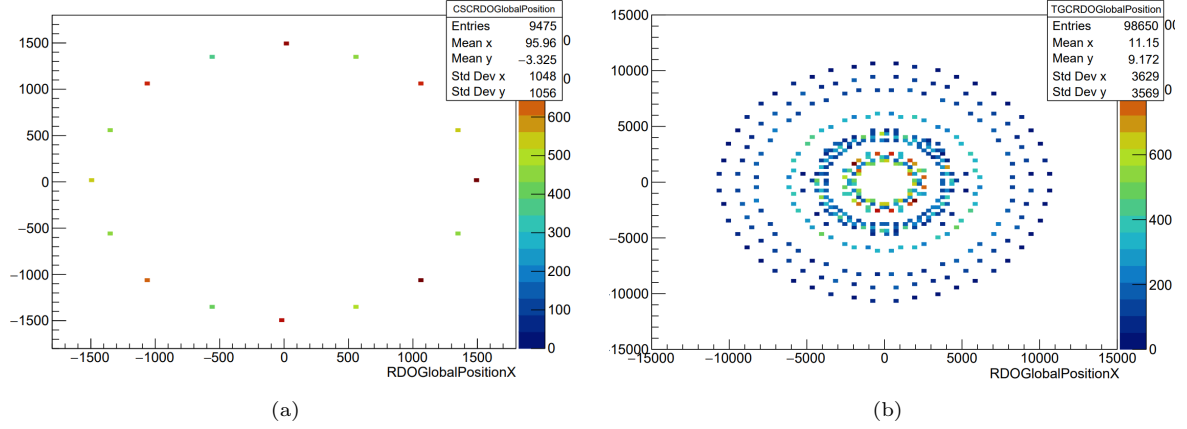


Figure 5.9: The x-y plane view of the (a) CSC and (b) TGC for the RDO outputs, both were generated with the 4000 $Z \rightarrow \mu\mu$ events which were accessed on 12th October, 2021.

into drift distance with associated uncertainty, and strips with charge and time information are transformed into clusters. The identifiers of the PRDs are assigned offline, as they do not carry information from the electronic channels. Each PRD has an associated pointer that provides access to the position information in each readout element.

The output of the reconstruction process consists of the Analysis Object Data (AOD) and Event Summary Data (ESD) classes. These classes contain information that can be used for physics analysis, as discussed in section 5.2.4. The muon validation algorithm utilizes the ESD from the reconstruction, as it provides more detailed information about the tracks and other objects used in the analysis. However, the AOD is more commonly accessed and used in physics analysis due to its ease of access.

Not all the physics processes need to be reconstructed with extremely high accuracy, the fast simulation, as shown in figure 5.5, aim at speeding up the part of the simulation process with little physics penalty. The fast simulation uses the outputs from the event generation and output the ESD or AOD directly, which sufficiently reduced the CPU time.

There are six reconstruction validation algorithms for each of the muon subdetector, the PRD outputs are monitored by the selected variables which are summarised in table 5.5. The isStrip variable is corresponding to the measure ϕ , it equal to 0 if the positions are measured by wires (measure $\phi = 0$), and it is 1 if the positions are measured by strips (measure $\phi = 1$). The reconstructed CSC and TGC Run-3 layout in term of the x-y plane view are shown in figure 5.10. The perfect shape of the subdetectors are clearly shown in term of PRD format, the CSC layout as shown in figure 4.12 (a) is simulated with a good modelling after the RDO-to-PRD transformation, a few pixels are lost in the TGC layout due to the low resolution.

5.3.4 Validation

In the muon validation of the ATLAS simulation chain, the outputs from the validation algorithms include Hits, SDOs, Digits, RDOs, and ESDs. The performance and plausibility of these transformation job outputs need to be monitored frequently, and any potential errors or problems from the simulation chain should be reported in detail. All these actions are expected to be carried out every day to prevent serious issues due to urgent breakdowns. The ATLAS Run

CSC PRD	TGC PRD
station η	
station ϕ	
global position X	
global position Y	
global position Z	
local position X	
chamber layer	gas gap
wire layer	isStrip
measure ϕ	channel
time	
charge	

Table 5.5: The list of the validation variables from the PRD outputs, which are monitored in the DCube-histogram monitoring system for the CSC and TGC subdetectors.

Time Testers (RTT) are used to automatically perform each test in the nightly build of the ATLAS offline software. The releases that require testing must pass several milestones before being declared ready for production.

The muon validation algorithms use 100 $Z \rightarrow \mu\mu$ events as input at the time of writing. These generated events serve as input for the simulation algorithm, and the resulting Hits are stored in a root file. Before the output Hits root file is input to the digitization validation algorithm, the selected variables for CSC and TGC listed in table 5.2 are read and compared to the results of the last nightly test. This comparison is performed using the DCube-histogram monitoring system to ensure result consistency. Further details of these comparisons are discussed in section 5.3.5.

The generated output Hits root file is used as input for the digitization algorithm, which transforms the Hits into RDOs. All the SDOs, Digits, and RDOs are saved in a single output root file following the Hits-to-RDO transformation. The selected variables listed in table 5.3 and table 5.4 for the CSC and TGC subdetectors, respectively, are also monitored using the DCube-histogram monitoring system.

In muon validation, the RDO-to-ESD transformation is the most crucial step. The reconstruction validation algorithm outputs the PRD variable histograms and saves them in the ESD root file. The selected variables listed in table 5.5 for the CSC and TGC are also subject to monitoring. Additionally, the reconstruction results are further validated by running the same reconstruction process using the *Athena multi-threaded* (AthenaMT) algorithm. AthenaMT is an algorithm [104] that enables the execution of multiple tasks concurrently within a single program. In other words, multiple events can be processed simultaneously. The outputs from the AthenaMT runs are expected to match the normal outputs. In the current muon validation algorithm, nightly runs of the reconstruction are scheduled with 1 thread and 5 threads. The results from the run with 5 threads are compared with the normal results as well as the results from the run with 1 thread. Any

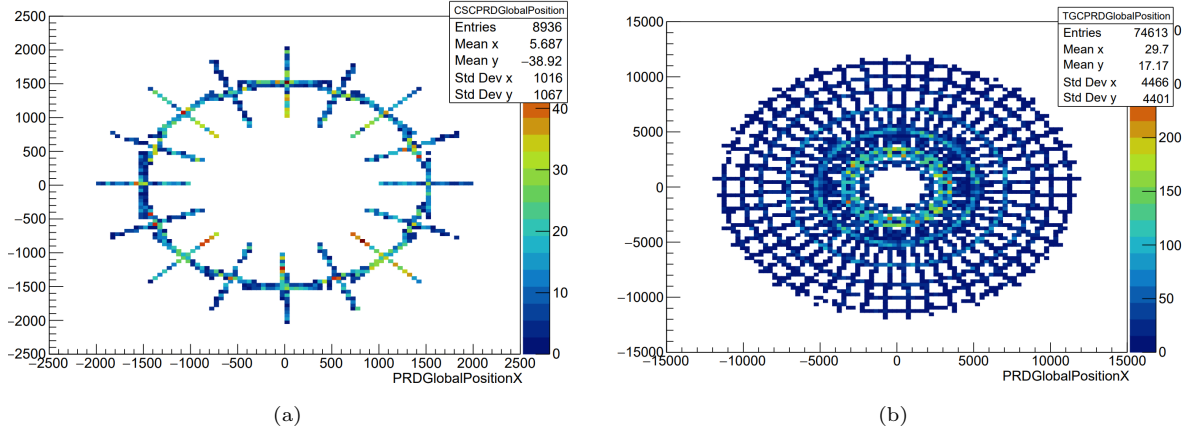


Figure 5.10: The x-y plane view of the (a) CSC and (b) TGC for the PRD outputs, both were generated with the 4000 $Z \rightarrow \mu\mu$ events which were accessed on 12th October, 2021.

differences observed in these comparisons will be reported. Figure 5.11 illustrates the complete validation chain for the Run-3 muon spectrometer layout.

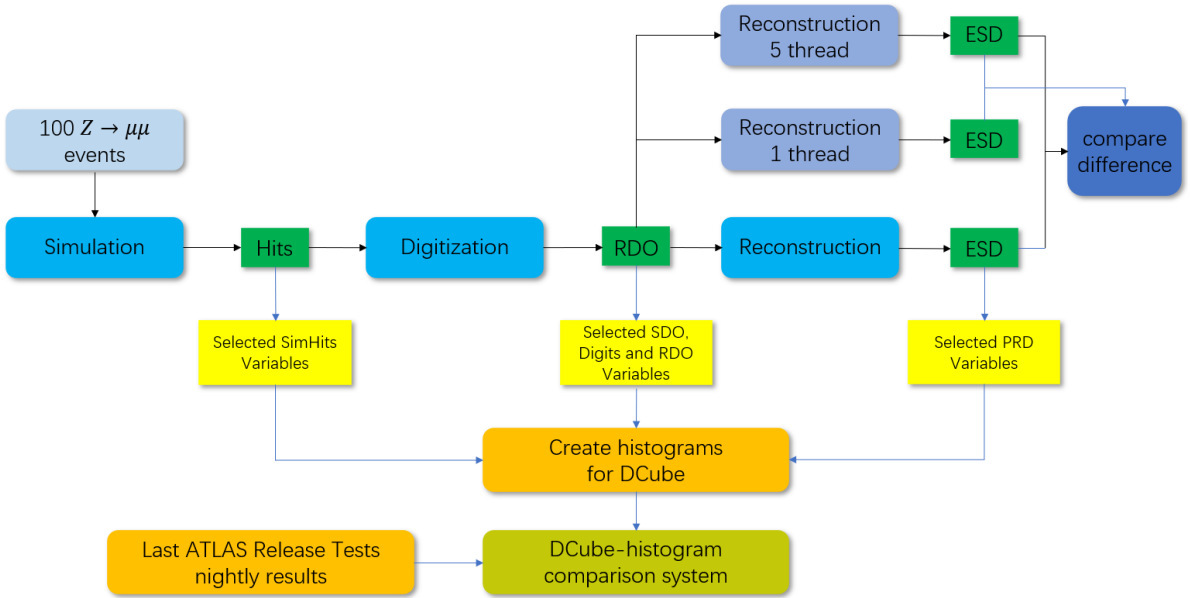


Figure 5.11: The schematic view of the muon validation full chain for Run-3 layout. The selected variables are listed in table 5.2, 5.3, 5.4 and table 5.5. The last ATLAS Release Tests nightly results refer to the DCube histograms from the previous results of the same chain.

5.3.5 Performance Monitoring

The selected validation variables from each transformation job outputs are monitored by comparing with the previous nightly outputs from the same validation chain, as shown in figure 5.11. The *DCube-monitoring system* [105] is employed for comparing the histograms with specific algorithms automatically. The DCube system can be used for performing the histogram-comparison jobs with a fixed reference or a previous nightly output, and it can be easily implemented with the

ATLAS Release Tests (ART) and compiled with the root files.

After each transformation, the histograms of the selected validation variables were retrieved from the output root files (Hits, RDO and PRD) and saved in another root file which will be used for the DCube comparison. The same set of histograms but produced from the previous nightly tests were downloaded and used as the reference in the comparisons. DCube system provide several statistical tests for the comparison of the histograms, they are listed as follow:

- Kolmogorov-Smirnov (KS) test [106]: The default statistical test of the DCube system, which is a common method for comparing a sample with a reference probability distribution.
- Pearson's χ^2 test [107]: The statistical test can be used to compare the observed distribution to the reference distribution, to test the goodness-of-fit of the reference distribution to the observed distribution.
- Bin-by-bin test: Compare the observed distribution to the reference distribution in a bin-by-bin approach.
- Mean value of y: Rather arbitrary test on the mean value of y (for profile plots only).

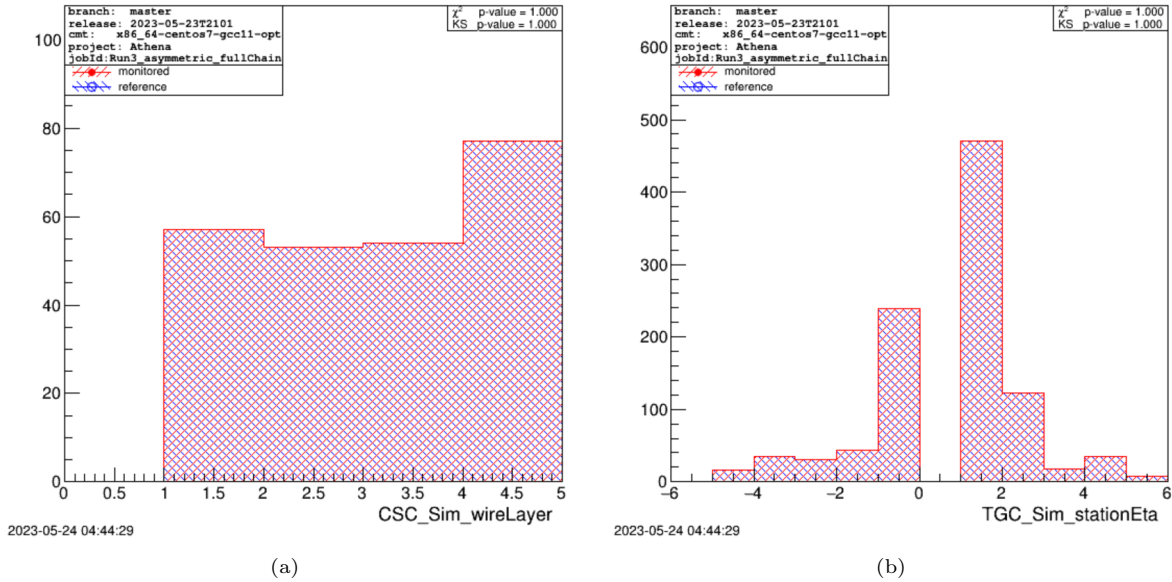


Figure 5.12: The example outcomes from the DCube-histogram monitoring system, generated at 24th May, 2023. The example comparisons for the validation variables of (a) wire layer for the CSC and (b) station η for the TGC from the Hits output root file.

The current muon validation used the KS test and the Pearson's χ^2 test in the DCube-histogram comparisons. In each of the DCube-histogram comparisons, a p-value is computed. Only those comparisons with p-value greater than 0.95 can pass the quality check, and warnings will be reported if the p-value of a comparison is within the range $0.75 < \text{p-value} < 0.95$. While failure will be reported if the p-value is lower than 0.75. Figure 5.12 shows two example outcomes from the DCube-histogram monitoring system, both example shows perfect match between the observed distribution and the reference distribution, the p-value is equal to 1.

The DCube-histogram monitoring system will compare all the information from the selected variables. However, in the case that a histogram is found to have a p-value lower than 0.75 (when a

failure is reported), actions are needed for locating the problematic positions. The breakdown of the station η and station ϕ can effectively help on finding where the reconstruction breaks. Therefore, the cuts on the station η and station ϕ are implied to the monitoring procedure; the negative and positive η cut can help on indicating the direction, while the ϕ sector cut can provide the performance of a specific sector. The η and ϕ cuts are also used for saving resources, as it is not necessary to save all the information.

Chapter 6

Object Reconstruction and Identification

Physics objects are reconstructed and identified before conducting any physics analysis. During the reconstruction process, bytestream data and simulated electronic signals are converted into the Raw Data Object (RDO) format. The observed characteristics of particles from proton-proton collisions within the detector are reconstructed and saved in the output.

In the Higgs LFV analysis, the physics objects relevant to the topology of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ events include electrons, muons, the visible products of hadronically decaying τ leptons ($\tau_{had-vis}$), jets, and missing transverse energy. These reconstructed objects are employed for selections and signal region categorization.

6.1 Tracks

The trajectories of the charged particles are reconstructed as tracks in the ATLAS detector. There are thousands of particles in each event which need to be reconstructed precisely and efficiently. The tracks information are used in many other objects reconstructions such as the primary vertex, leptons, pile-up removal and jet flavour tagging.

The trajectory of a charged particle in a magnetic field inside the detector is described by five track parameters, as shown in Figure 6.1: d_0 and z_0 represent the transverse and longitudinal impact parameters referring to the interaction points, the angular parameters ϕ and θ , and the charge-to-momentum ratio q/p . These parameters are used to define the space points from detector information, providing the three-dimensional position. Track reconstruction is performed in the inner detector using the inside-out algorithm, beginning with seed finding in the silicon layers. Seeds are combinations of space points that help reduce fake tracks. Each track candidate is assigned a score based on its track parameters and hit topology, including the number of shared hits and hits on the track. Those with poor quality are rejected to resolve ambiguity issues. The final step involves extending the silicon tracks into the TRT and performing a track fit to obtain the final track parameters. The outside-in algorithm is also employed to reconstruct tracks for particles produced far from the beamline, primarily electron candidates from photon conversion.

Track reconstruction in Run-2 experiments has achieved impressive performance with sufficiently high efficiencies and a low fake rate of around 0.1% [109]. To address the challenges posed by Run-3

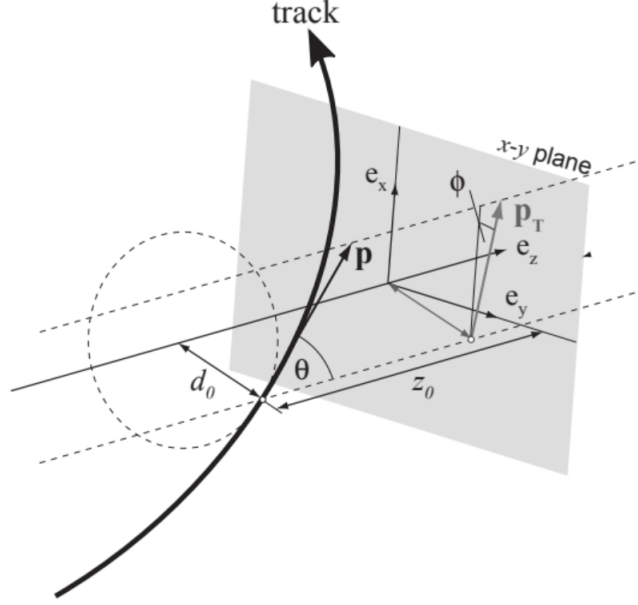


Figure 6.1: Track of a charged particle in the ATLAS right-handed coordinate system [108].

and HL-LHC experiments, upgrades have been made to the software [110] and the Fast Tracker (FTK).

6.2 Primary Vertex

The interaction point represents the designated position where particles collide. However, the actual location of interactions may differ from this design position. The primary vertex is defined as the true interaction point resulting from the particle collision. Primary vertex reconstruction employs reconstructed tracks that are selected based on specific criteria. If there are multiple primary vertices reconstructed within an event, the vertex with the highest sum of p_T^2 for the associated tracks is chosen.

For the vertex reconstruction, the reconstructed tracks must have momenta $p_T > 400$ MeV and fall within the range of $|\eta| < 2.5$. The silicon layers require a minimum of 9 hits (11 hits) if $|\eta| < 1.65$ ($|\eta| > 1.65$), and there should be at least one hit in the first two pixel layers [111]. To mitigate the fake rate, the pixel detector should have no holes, and the SCT can have at most one hole. A hole refers to an expected measurement on the detector surface that goes unobserved.

There are two stages of the vertex reconstruction: the vertex finding and vertex fitting. The vertex reconstruction is an iterative procedure between these two stages, the details of the procedure is shown as follow:

1. Select the tracks which passed the criteria defined above.
2. Find the seed position for the first selected vertex by considering the impact parameter z_0 of the track with respect to the beam spot, the first seed position is taken as the centre of the beam spot.
3. Using the track and the seed to fit the best vertex position. Each track is weighted or scored, those less compatible tracks are assigned lower score and iterated to search for the best vertex

position.

4. After the vertex position is determined, those tracks with lower score are taken away for another iteration for searching another vertex.
5. Two associated tracks are required for each vertex. Repeat above iteration until all associated tracks are used and all the vertices have been found.

This algorithm produces a set of vertex positions and their covariance matrices, the width and position of the reconstructed beam spot are used in the fit of the primary vertex as additional measurement.

In the Run-3 experiments, the large increase in the luminosity poses a challenge for the primary vertex reconstruction performance. In order to meet this challenge, new ATLAS tools have been developed: a Gaussian track density seed finder and an adaptive multi-vertex finder, the details are introduced in [112]. In addition, the primary vertex finder based on a hybrid deep-learning algorithm is also developed [113], the new vertex finding algorithm use the tracks as the inputs of a neural network and outputs the predictions for the position of primary vertex.

6.3 Electrons

Electrons are reconstructed from energy deposits in the electromagnetic calorimeter which are matched to a track reconstructed in the inner detector. The trajectory of an electron throughout the pixel detector, SCT and TRT before it was stopped in the electromagnetic calorimeter. In the mean time, a photon is induced by the interaction of the electron with the material in the tracking detector. Figure 6.2 illustrates the schematic diagram of the path of the electron and the induced photon.

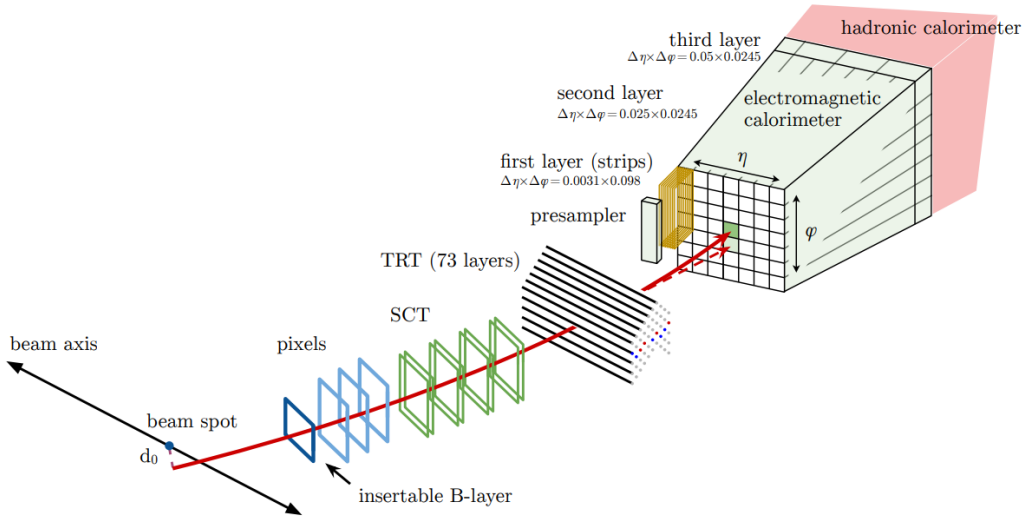


Figure 6.2: The schematic diagram of the trajectory of an electron pass through the detector, the path of the induced photon is shown in dash line [114].

The electron reconstruction is carried out within the precision region of the ATLAS detector with $|\eta| < 2.47$. As shown in Figure 6.2, the $\eta \times \phi$ space of the calorimeter layer is divided into a grid of 200×256 elements with a size of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ [114]. The energy from the

first three layers of the electromagnetic calorimeter and the presampler is grouped to form the energy of the tower. An algorithm is used to seed the energy cluster candidates by employing towers with variable sizes [115], which is known as the *sliding-window* algorithm [116].¹ The reconstructed electron consists of a *supercluster* and a matched track. The supercluster is a cluster built from energy deposits in the calorimeter and reconstructed using an algorithm with *satellite clusters*, defined as clusters that fall within a window around the seed cluster barycenter [115].

The next step is the track reconstruction, following the procedure introduced in section 6.1. Additionally, an optimized Gaussian-sum filter (GSF) method [117] is employed to improve the measurement of electron energy loss in material. The energy loss due to bremsstrahlung is also measured using the GSF method. For primary electron track reconstruction, the *egamma* working group in ATLAS recommends including a requirement on the impact parameters: $\sigma(d_0)/d_0 < 5.0$ and $z_0 \times \sin(\theta) < 5.0$.

The electron and photon reconstruction procedure begins with preparing the tracks and clusters discussed above. These tracks and photon conversion vertices are then matched to the *topo-clusters*, which are the measured clusters of energy in electromagnetic and hadronic calorimeter cells connected topologically. The reconstruction algorithm for topo-clusters is discussed in [118]. The track-matched topo-clusters are used to seed electron superclusters, while photon superclusters are seeded from both track-matched topo-clusters and photon-conversion-matched topo-clusters; electron and photon superclusters are built separately. Additional requirements are applied to an electron supercluster seed: the transverse energy must be greater than 1 GeV, and there should be at least four hits in the silicon layers. The photon supercluster seed is required to have a transverse energy greater than 1.5 GeV [115]. After calibrating the energy and correcting the position of the selected superclusters, the tracks are matched to the electron superclusters, and the photon-conversion vertices are matched to the photon superclusters. Both types of superclusters are then used to build the analysis objects.

The superclusters of electrons and photons are built separately. However, a seed cluster can produce both an electron and a photon, necessitating an ambiguity resolution procedure to identify whether it is an electron or a photon. Any specific requirements for analysis can be applied in the final classifications of electrons and photons. The energies of electrons and photons are recalibrated based on the algorithm introduced in [119].

The last step is to identify electrons using a Likelihood-based method. The signal in the likelihood function is the prompt electron, while the backgrounds consist of misidentified events, such as jets mimicking prompt electrons, electrons from photon conversions, or non-prompt electrons from heavy-flavour hadron decays. Full details are provided in [115]. The likelihood discriminant has four different criteria defining four operating points: VeryLoose, Loose, Medium, and Tight, which correspond to increasing thresholds for the likelihood discriminant. Different operating points are chosen based on analysis requirements for signal efficiency and background rejection. For example, considering the identification of a prompt electron with $E_T = 40$ GeV, the corresponding identification efficiencies are 93%, 88%, and 80% for Loose, Medium, and Tight, respectively [115].

In addition to electron identification, distinguishing prompt electrons in signal processes from background processes, such as heavy quark decays and hadrons misidentified as electrons, is necessary. The true electron signature can be represented by the activity surrounding the candidate object in an area with a radius of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. This activity can be quantified by nearby

¹The algorithm with the variable-size feature is an improved version; the fixed-size version was used in the analysis with 2015-2016 data previously [114].

charged particle tracks or energy deposits in the calorimeters, resulting in two types of isolation variables: track-based isolation and calorimeter-based isolation. The isolation criteria for electrons are analysis-specific, with details summarized in chapter 8. Four isolation working points are defined: Gradient, HighPtCaloOnly, Loose, and Tight; details can be found in [115]. Calorimeter-based isolation depends on cell resolution, requiring additional corrections due to particle energy leakage outside the activity area. For track-based isolation, tracks not used in the vertex fit should satisfy the condition $|z_0| \sin \theta < 3 \text{ mm}$ [115].

6.4 Muons

Muons are reconstructed based on the information from the tracking detectors in the muon spectrometer and inner detector, by combining the tracks that have consistent trajectories and curvatures. Information from the calorimeters is also used for the cases of large energy loss in the calorimeters by determining the impact parameters.

The reconstructed tracks in the inner detector are discussed in section 6.1 and used to combine with the reconstructed muon spectrometer stand-alone tracks. The tracks reconstruction in the muon spectrometer are short straight-line segments that are reconstructed from the hit patterns inside each muon subsystem. The parabolic trajectory of the muon in the magnetic field is used to construct the preliminary track candidates by combining the segments in different stations [120], the three-dimensional track candidates are constructed by combining with the information from the precision measurements of the second coordinate from the trigger chambers. The curvature of the muons in the magnetic field are further analysed by performing χ^2 fit, reducing other noises in the measurements [120]. The refit is performed until all accepted track candidates satisfied the selection criteria.

The global muon reconstruction uses the information provided by the inner detector, calorimeters and muon spectrometer. Five algorithms are developed depending on which subdetectors are dominant in the reconstruction procedure [120]:

- Combined (CB) muon: Tracks are reconstructed independently in the inner detector and the muon spectrometer by using the procedures discussed above. The hits from the muon spectrometer and the inner detector are used to fit a combined track, and the energy loss in the calorimeters are taken into account. The hits with the associated tracks in the muon spectrometer are fitted in the stand-alone reconstruction procedure, they maybe refitted based on the particle trajectory from the combined fit. In forward region with $|\eta| > 2.5$, the combined muon is constructed by combining the tracks in the muon spectrometer and the segments reconstructed from hits in the pixel and SCT detectors.
- Inside-out (IO) combined muon: Large majority of muons are reconstructed based on the outside-in algorithm, in which the muons are first reconstructed in the muon spectrometer and then extrapolated inward and matched to a track in the inner detector. The inside-out algorithm serves as complementary way to retrieve the loss in the reconstruction efficiency due to the dependence on an independently reconstructed track in the muon spectrometer. The IO muons are reconstructed by extrapolating the inner detector tracks to the muon spectrometer with an additional requirement that at least three loosely-aligned hits in the muon spectrometer, and the energy loss is also taking into account. Such extrapolation algorithm can cover the

deficiencies of the muon spectrometer such as its limited coverage and detection of the muons with low p_T .

- Muon-spectrometer extrapolated (ME) muons: The muon spectrometer has a coverage of $|\eta| < 2.7$, while the inner detector only covers the range $|\eta| < 2.5$. The gap between them are filled by the ME muons, whose parameters are extrapolated to the beamline.
- Segmented-tagged (ST) muons: This type of muons are reconstructed rely on the tracks in the inner detector which are extrapolated to the muon spectrometer with an additional tighter requirement on the angular matching with at least one reconstructed muon spectrometer segment.
- Calorimeter-tagged (CT) muons: The minimum-ionising particles are main signature for muon identification, the tracks in the inner detector are extrapolated to the calorimeters, the muons are those can be matched an energy deposit in the calorimeters compatible with a minimum-ionizing particle. The CT muons are required to have $p_T > 5$ GeV in order to avoid overlapping with the ST muons which also use the ID tracks directly.

In the Higgs LFV analysis, large majority of the muons belong to the CB type.

Muon identification procedure is used to further select the high quality muon candidates with additional requirements and to reject the muons from the background processes such as pion and kaon decays. Similar to the electron identification, different selection criteria are implied upon analyses. There are three identification working points: *Loose*, *Medium* and *Tight*, and two special working points - the high- p_T and the low- p_T are designed for the analyses focus on the extreme phase space regions. Only the CB and IO muons are accepted in the *Medium* working point within a coverage of $|\eta| < 2.5$ in the inner detector, with additional requirement on the q/p compatibility [120]. All the muons passing the *Medium* working point are accepted by the *Loose* working point, with an additional acceptance for the CT and ST muons in the range $|\eta| < 0.1$ for covering the loss in the CB muon reconstruction efficiency, more detail selection criteria is shown in [120]. Only the CB and IO muons with at least two precision stations are accepted for the *tight* working point.

The muon isolation has a large rejection power for the non-prompt muons from the heavy-flavour hadronic decays, and it is a similar procedure compare to the electron isolation. Different isolation working points requirements are introduced in [120], which include track-based isolation and calorimeter-based isolation. In the Higgs LFV analysis, the used working point is FCTIGHT_FIXEDRAD, which requires $E_T^{topocone20}/p_T^\mu < 0.15$ for the calorimeter-based isolation, where $E_T^{topocone20}$ is defined as the sum of transverse energy of topological cell clusters in a cone with size $\Delta R = 0.2$ surrounding the muon candidate [120]. The track-based isolation requires $p_T^{varcone30}/p_T^\mu < 0.04$, where $p_T^{varcone30}$ is defined as the scalar sum of the transverse momentum of the tracks in the inner detector associated with the primary vertex in an $\eta - \phi$ cone with size $\Delta R = \min(10\text{GeV}/p_T^\mu, 0.3)$ surrounding the muon candidate [120].

6.5 Hadronic Taus

Two decay modes of the tau lepton are considered in this analysis: the leptonic decays $\tau \rightarrow \ell + \nu_\ell + \nu_\tau$ and the hadronic decays $\tau \rightarrow \text{hadrons} + \nu_\tau$. However, the leptonic taus (τ_ℓ) are almost indistinguishable from prompt electrons or muons; the main difference is a displacement

of the tau decay vertex with respect to the primary vertex. Among the decay products of the hadronic tau, the neutrino can penetrate through the detector without being detected, and the visible products are mostly one or three charged pions and up to two neutral pions.

The reconstruction algorithm for these $\tau_{had-vis}$ objects is initiated by jets reconstructed using the anti- k_t algorithm [121], with a distance parameter of $R = 0.4$ [122]. The inputs to the jet algorithm are the calorimeter TopoClusters calibrated using local hadronic calibration, and a jet is required to have $p_T > 10$ GeV and be within the coverage $|\eta| < 2.5$. Details of the jet reconstruction are introduced in section 6.6.

The identification of the primary vertex and tau lepton production vertex uses different algorithms. The reconstructed primary vertex in an event is required to have at least three associated tracks. For events with more than one primary vertex candidate, the one with the highest sum of P^2T of the tracks is used. To reduce the pile-up effects on the reconstruction efficiency, an algorithm associated with the tau vertex is developed based on the tau candidate tracks with $p_T > 0.1$ GeV. The p_T of the tracks in the area with $\Delta R < 0.2$ around the jet seed is summed, and the primary vertex candidate with the highest sum is chosen as the tau vertex.

The tracks that satisfy the mentioned conditions must further fulfill the following criteria: at least seven hits in the pixel and SCT layers and two associated hits in the pixel layers, $|d_0| < 1.0$ mm, and $|z_0 \sin \theta| < 1.5$ mm (with respect to the tau vertex) [122]. Selected tracks are used to optimize the fraction of tau decays with 1 charged pion and with three charged pions, reconstructed with the correct charged particle multiplicity.

To identify the products of the τ_{had} decays, a reconstruction algorithm for the neutral pion is also developed. This algorithm employs the machine learning technique via the BDT method to classify the clusters used in a neutral pion candidate. Kinematic variables of the tracks and the clusters are used as the BDT inputs; details are provided in [120].

Energy calibration is performed on the $\tau_{had-vis}$ candidates to correct the energy deposits in the calorimeters to the average energy carried by the decay products. Details of the energy calibration of the $\tau_{had-vis}$ candidates are provided in [123]. Two calibration methods are employed: the baseline calibration, which provides high energy resolution for the phase space region with high p_T , and the Boosted Regression Tree (BRT) based calibration, which corrects for differences between data and simulated events.

The tau reconstruction algorithm described above provides extremely low discrimination power in separating the reconstructed $\tau_{had-vis}$ initiations from hadronic tau decays from jets initiated by quarks or gluons. A Recurrent Neural Network (RNN) method [124] is employed to identify the $\tau_{had-vis}$ objects from the jets; details of the training strategies are introduced in [124]. Both low-level and high-level variables are used as inputs for the RNN training, such as the tracks and clusters of energy in the calorimeters associated with the $\tau_{had-vis}$ candidates. Four working points are determined based on the RNN score, in increasing order of background rejection: VERYLOOSE, LOOSE, MEDIUM, and TIGHT. Background rejection and signal efficiency are measured separately for taus with single-track (1-prong) and taus with three tracks (3-prong); table 6.1 shows the corresponding identification working point criteria.

Electrons have very similar signature of the $\tau_{had-vis}$ candidates with 1-track and could be misidentified, and most of these background contributions come from the $Z \rightarrow ee + \text{jets}$ processes in the Higgs LFV analysis. Another multivariate discriminant via BDT is used to reject those electrons which misidentified as the $\tau_{had-vis}$ candidates, information from the calorimeter and

Working Point	RNN Score		Signal Efficiency	
	1-prong	3-prong	1-prong	3-prong
VERYLOOSE	> 0.05	> 0.05	95%	95%
LOOSE	> 0.15	> 0.25	85%	75%
MEDIUM	> 0.25	> 0.30	75%	60%
TIGHT	> 0.40	> 0.55	60%	45%

Table 6.1: The respective RNN score and signal efficiency in each of the four working point.

tracking detector are used as the inputs of the eBDT, the transition radiation information from the TRT system has the largest discrimination power among others. Three working points are also developed based on the BDT score: *loose*, *medium* and *tight*, which yield signal efficiencies of 95%, 85% and 75%, respectively.

Muons with large energy deposits in the calorimeters and the track could be misidentified as the $\tau_{had-vis}$ candidates, but those muon faking $\tau_{had-vis}$ candidates can be rejected with very high efficiency by using a simple cut-based analysis.

6.5.1 New Algorithms used in Run-3

The reconstruction, identification and energy calibration of the hadronic decaying τ leptons have been improved for the ATLAS Run-3 experiments and reprocessed for the Run-2 data, the details and the performance of the updates are introduced in [125].

First of all, the tau vertex identification algorithm has been updated for meeting the challenge of much higher pile-up environment. A different method based on the output of the Adaptive Multi-Vertex Reconstruction algorithm [125, 126] is used in the Run-3 experiments, the efficiency of identify the correct hadronic τ production vertex using this algorithm is even higher compare to the primary vertex.

Secondly, the identification algorithm for the tracks associated with the τ lepton production vertex is also updated, a RNN-based discriminant is used in Run-3 instead of the BDT used in Run-2 experiments. The RNN method provides a flexible architecture which allows the tracks to be analysed dynamically as an input of variable length per $\tau_{had-vis}$ candidate. Large improvement is achieved by using the new RNN classification, the average efficiency is increased from 70% to 90% for the 1-prong candidates and is increased from 65% to 75% for the 3-prong candidates [125].

The RNN τ lepton identification algorithm discussed above for separating the truth $\tau_{had-vis}$ from the misidentified candidates originating from the jets is kept in the Run-3 experiments. The RNNs have been retrained in the same way using the new algorithms, except that three of the input variables: $p_T^{seed-jet}$, d_0^{track} and $z_0^{track} \sin\theta$ have been replaced by $p_T^{\tau_{had-vis}}$, d_0^{TJVA} and $z_0^{TJVA} \sin\theta$ [125], respectively, where TJVA means tau production vertex association.

The eBDT used for rejecting the electron misidentified as $\tau_{had-vis}$ candidates have has been replaced by a new approach based on RNN classification in the Run-3 experiments. Information from the tracks and calorimeters are used as inputs to the RNN trainings, details of the training strategies are introduced in [125]. The new RNN classification has a better performance with a factor of three on the background rejection [125] compare to the old eBDT algorithm.

There are five major decay modes of the τ_{had} , which either with one charged hadrons ($\pi^\pm$ or $K^\pm$) or three charged hadrons, and either with zero, one or more neutral pions. The decay mode classification algorithm used in previous Run-2 experiments is BDT-based [127], a new algorithm based on the DeepSet Neural Network is developed [125] for improving the classification power of the τ_{had} decay modes, which also uses to calculate the four-momentum of the $\tau_{had-vis}$ candidates and significantly improve the energy calibration. The details of the training strategies are introduced in [125], and the DeepSet NN improves the diagonal efficiency by 9% compare to the BDT-based approach.

6.6 Jets

Jets are produced through the hadronization of a quark or gluon, forming a narrow cone that contains hadrons. In ATLAS experiments, jets are identified using the anti- k_t algorithm with a distance parameter $R = 0.4$, applied to particle flow objects [128]. The inputs for both the particle flow and anti- k_t algorithms are selected tracks in the inner detector and energy deposits in the calorimeter. While jets reconstructed from tracks are constrained by the tracking detector's coverage, most jets are reconstructed primarily based on calorimeter energy deposits, known as calorimeter jets.

The particle flow algorithm combines tracks and calorimeter energy deposits to reconstruct hadronic jets. In ATLAS experiments, jets predominantly consist of charged pions, although other constituents can include neutral hadrons and photons from neutral hadron decays. Thus, jets are reconstructed from a combination of particle flow objects. The particle flow algorithm, introduced in [128], matches tracks to each TopoCluster, and employs an energy subtraction procedure to resolve overlaps between momentum measurements in the inner detector and energy measurements in the calorimeters.

Truth jets are required to have a lifetime greater than 10 mm to exclude muons, neutrinos, and particles from pile-up activity [129]. These truth jets are measured at the particle-level energy scale. For jet energy calibration, truth jets with $p_T > 7$ GeV and $|\eta| < 4.5$ are employed in the calibration procedure using MC simulation [129]. The jet energy scale (JES) calibration, outlined in [128, 129], corrects reconstructed jet energies to match the average values of truth jet energies. Additional corrections are applied through the *global sequential correction* (GSC) [130], addressing response fluctuations caused by the originating parton's flavor and the composition of hadrons within the jet fragmentation [128]. In situ JES calibration is performed to rectify differences in jet response between data and MC simulation.

To differentiate high- p_T jets from misidentified jets due to non-collision backgrounds, cleaning criteria are utilized [131]. Only jets with $p_T > 20$ GeV and $|\eta| < 4.5$ are considered in the Higgs LFV analysis. Hadronic jets appear as groups of topologically linked energy deposits in the calorimeters, making pile-up effects substantial for measured jet energy and structure. To mitigate these effects, a *jet-vertex tagger* (JVT) algorithm [132] is applied to jets with $p_T < 60$ GeV and $|\eta| < 2.5$. Pile-up jets in the forward region are suppressed using a forward JVT algorithm [133], which leverages jet geometry and topological correlations in pile-up interactions, applying the criteria $p_T < 60$ GeV and $|\eta| > 2.5$.

Top quarks predominantly decay into bottom quarks, contributing to background in the Higgs LFV analysis. Accurate identification and rejection of jets initiated by b -quarks are crucial. This is

achieved through the DL1r b -tagging algorithm [134], exploiting the distinctive characteristics of b -hadrons, such as long lifetime, high mass, and multiplicity of decay products, along with properties of heavy-quark fragmentation. The algorithm’s outputs are utilized for all reconstructed jets with $p_T > 20$ GeV and $|\eta| < 2.5$. The fixed 85% efficiency working point is employed in the Higgs LFV analysis.

6.7 Overlap Removal

The reconstructed objects used in the Higgs LFV analysis are not built from a set of mutually exclusive tracks or calorimeter clusters, therefore, the situations that two different objects share most of their constituents are possible and must be considered. The ambiguity resolution is done by an overlap removal (OLR) procedure which is applied to the *baseline* objects discussed in this section, by using the ASSOCIATIONUTILS package. The STANDARD working point is used in the Higgs LFV analysis based on the criteria listed in table 6.2, the overlap removal procedure is proceeded in the order specified in the table, and the electron has the highest priority.

Step	Reject	Against	Criteria
1	electron	electron	shared track, $p_T^1 < p_T^2$
2	$\tau_{had-vis}$	electron	$\Delta R < 0.2$
3	$\tau_{had-vis}$	muon	$\Delta R < 0.2$
4	muon	electron	is-calo muon and shared ID track
5	electron	muon	shared ID track
6	jet	electron	$\Delta R < 0.2$
7	electron	jet	$\Delta R < 0.4$
8	jet	muon	$N_{tracks} < 3$ and ghost-associated or $\Delta R < 0.2$
9	muon	jet	$\Delta R < 0.4$
10	jet	$\tau_{had-vis}$	$\Delta R < 0.2$

Table 6.2: The overlap removal procedure with STANDARD working selection, the selection criteria are applied sequentially in order.

6.8 Missing Transverse Energy

The missing transverse energy (E_T^{miss}) is an estimate of the imbalance in the transverse momentum in the detector, which is an experimental proxy for the transverse momentum carried by the particles which can penetrate through the detector without being detected, those particles could be the SM candidates of neutrinos and the BSM candidates of the undetected new particles.

The reconstructed missing transverse energy includes two contributions, the first contributions are the hard-event signals come from the reconstructed electrons, muons, τ leptons and photons. The

second contributions are the soft-event signals comprising the detector signals which are recorded for one triggered event but not used by the reconstructed objects listed above [135]. The E_T^{miss} is calculated as the negative vector sum of the transverse momenta of all the reconstructed final-state objects and the unused tracks:

$$\mathbf{E}_T^{miss} = -\Sigma \mathbf{p}_T^e - \Sigma \mathbf{p}_T^\gamma - \Sigma \mathbf{p}_T^{\tau_{had}} - \Sigma \mathbf{p}_T^\mu - \Sigma \mathbf{p}_T^{jet} - \Sigma \mathbf{p}_T^{track} \quad (6.1)$$

The Higgs LFV analysis used the default TIGHT criteria of the official ATLAS missing transverse energy tool [135], the implemented criteria ensures that the forward jets with $|\eta| > 2.5$ used for the E_T^{miss} calculation have $p_T > 30$ GeV, and provide the ambiguity resolution. Dedicated OLR procedure is implemented in the E_T^{miss} reconstruction.

Chapter 7

Analysis Strategies

In this thesis, the primary objective of the Higgs LFV analysis is to search for observable instances of lepton flavor violation (LFV) decays involving the Higgs boson. The preparation of both data and Monte Carlo (MC) events employed in this analysis has been outlined. This chapter delves into the strategies implemented to detect LFV signals within the data, while comprehensive details regarding each step of the analysis will be provided throughout the subsequent sections of this thesis.

In this analysis, the searches focus on two Higgs LFV decays: $H \rightarrow \mu\tau$ and $H \rightarrow e\tau$. The τ -lepton can decay either hadronically as τ_{had} or leptonically as τ_μ and τ_e . Therefore, the final states of the searches are: $H \rightarrow \ell\tau_{had}$ (lephad) and $H \rightarrow \ell\tau_{\ell'}$ (lelep), where ℓ represents μ and e , and ℓ' is the light lepton with different flavour of ℓ . The schematic view of those decays are illustrated in figure 7.1. About 65% of the total τ decays are hadronic decays and the rest are leptonic decays. In the lelep channel, the possible Higgs LFV signals with final states which include two leptons have the same flavour ($e\tau_e$ and $\mu\tau_\mu$) are not considered due to the overwhelming background of $Z \rightarrow \ell\ell$, but they could be the potential signals in the Run-3 and HL-LHC experiments.

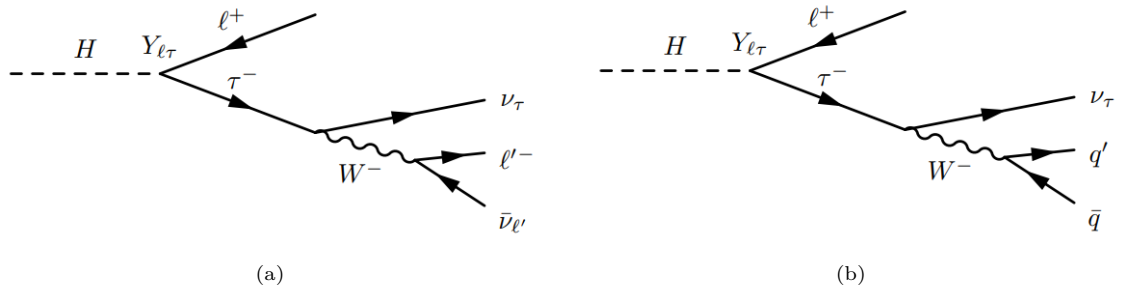


Figure 7.1: The schematic view of the LFV decays of the Higgs boson for the (a) $\ell\tau_{\ell'}$ and (b) $\ell\tau_{had}$ final states. The $Y_{\ell\tau}$ indicates the off-diagonal Yukawa coupling term [51].

Two independent background estimation techniques are exploited in this analysis: the MC-template method and the symmetry method. In the Higgs LFV analysis with Run-2 data [51], both lelep and lephad channel is analysed in the MC-template method. However, only the lelep channel is analysed by using the symmetry method. Most of the research works in this thesis are focused on the $\ell\tau_{had}$ final states using MC-template method, part of the contributions are also given to the analysis of the lephad channel using symmetry method.

7.1 MC-template Method

The MC-template method uses templates from MC simulations, while data is used to normalize the two main backgrounds. The misidentified backgrounds are estimated using data-driven methods. This approach is directly sensitive to the individual branching ratios of the LFV decays of the Higgs boson.

The events being analysed are further categorized based on clear signatures in the signal region. To reduce some of the major backgrounds, the corresponding event spaces are cut based on related kinematic variables. Additionally, dedicated control regions are defined for background normalization and to constrain systematic uncertainties. Details are discussed in chapter 8.

The MC-template method employs both MC simulations and data-driven techniques to estimate background contributions. The background processes are estimated from simulation, but dedicated control regions enriched with the major backgrounds are defined from data and used to determine normalization. The final state objects could be mimicked by other SM processes. In the $lele\mu$ channel, fake lepton contributions are estimated from data using a data-driven method known as the *ABCD* method, and objects misidentified as τ_{had} are estimated via the *Fake Factor* method [136]. Details are provided in chapter 10.

Kinematic variables with the largest classification power between the signal and background are used to improve the agreement, and an analysis based on a single variable is known as a *cut-based analysis*. Details are introduced in section 7.3, and mass-related variables are found to have the best performance in the cut-based analysis. To optimize signal-to-background separation, a multivariate analysis is performed based on a machine learning technique known as the Boosted Decision Tree (BDT) [137]. Each event is assigned a BDT score; events more likely to be a signal (background) have higher (lower) BDT scores. More details on training strategies and performance optimization are provided in chapter 11.

The systematic uncertainties considered in this analysis are introduced in chapter 12. The uncertainties that arise in this analysis can be grouped into three types: experimental uncertainties, and theoretical uncertainties for both signal and backgrounds. These uncertainties are used in the statistical analysis and are presented in the resulting measurements.

The BDT score distributions are used in the statistical analysis, which is discussed in chapter 13. The statistical analysis of the data uses a binned likelihood function constructed as a product of Poisson probability terms over all bins considered in the search. In this search, a simultaneous fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ branching ratios in the $e\tau$ and $\mu\tau$ channels is performed. If a significant excess of the LFV signal is observed, it can be claimed as a discovery of new physics depending on the sigma level. Otherwise, upper limits on the branching ratios are measured based on the data. The final results of the Run-2 analysis are presented in chapter 14.

7.2 Symmetry Method

The analysis employing the symmetry method [5] has a procedure very similar to the MC-template method discussed above. The main difference is that the symmetry method uses a data-driven analysis strategy (lepton-lepton channel only) and is more sensitive to the branching ratio difference between $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$. In other words, instead of directly targeting the individual branching ratios, the searches focus on measuring $Br(H \rightarrow e\tau) - Br(H \rightarrow \mu\tau)$ or

$Br(H \rightarrow \mu\tau) - Br(H \rightarrow e\tau)$, assuming that the branching ratio of the latter decay difference is zero. This assumption simplifies the measurements of observed (expected) branching ratios, upper limits, and significances.

The symmetry method exploits the symmetry of the prompt-lepton backgrounds in the SM under the exchange of electrons and muons. In other words, it relies on the assumption that the kinematic distributions of prompt electrons and prompt muons are approximately the same. Any signals of lepton-flavor-violating (LFV) decays of the Higgs boson break this symmetry.

To illustrate the presentation of the $H \rightarrow \mu\tau$ signal (as an example) within the symmetry method, events are simulated in a collinear mass distribution and must pass the baseline selection criteria. These simulated events include an added $H \rightarrow \mu\tau$ signal with a branching ratio of 10%, and a comparison between events with a leading muon and a subleading electron from the SM $\mu\tau + \mu\tau(Br(10\%))$ to events with a leading electron and a subleading muon, with applied efficiency corrections, is shown in Figure 7.2. The bump in the distribution represents the asymmetry and an LFV Higgs signal.

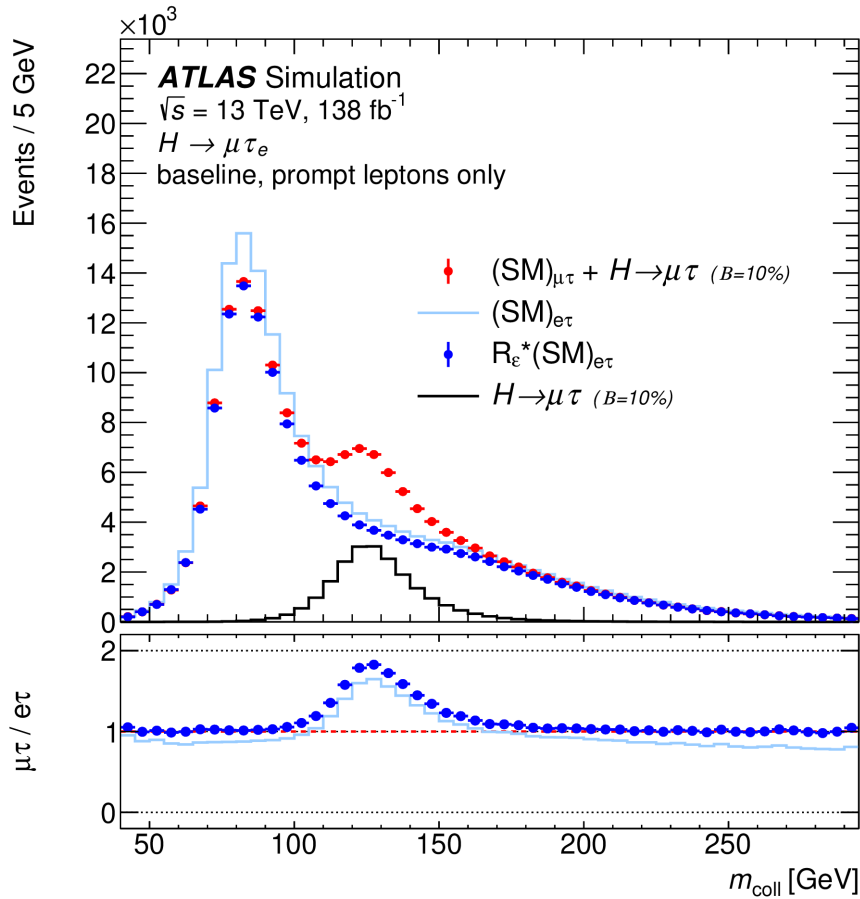


Figure 7.2: The collinear mass distribution with simulated events which pass the baseline selection criteria. The bottom pad shows the ratio between the $e\tau$ and $\mu\tau$ predictions. The blue points show the events with efficiency correction applied ($R^*(\text{SM})_{e\tau}$) [51].

The events are categorized following the same strategies as the MC-template method. For the $\ell\tau_{\ell'}$ final states, the event selection used in the symmetry method is almost the same as the one used in the MC-template method. For the $\ell\tau_{had}$ final states, the event selection used in the

symmetry method should be the same in both the $e\tau$ and $\mu\tau$ channels. The event selection criteria are modified based on the criteria used in the MC-template method and the $e - \mu$ symmetry.

To validate the $e - \mu$ symmetry, other effects that could invalidate it need to be taken into account. Objects mimicking prompt leptons should be rejected; misidentified leptons mainly originate from jets or heavy-flavour mesons. For the $\ell\tau_{had}$ final states, the main contributions come from jets imitating τ -leptons. The misidentified leptons contribute differently to the $e\tau$ and $\mu\tau$ final states, and their kinematic distributions are not symmetric. However, events containing misidentified τ -leptons are expected to be symmetric. Another effect could arise from the detector level; electrons and muons have different trigger and detection efficiencies (reconstruction, identification, and isolation). Scale factors must be estimated and applied to scale the final discriminant distribution to account for this asymmetry. The details of the efficiency correction are provided in Chapter 9.

In the cut-based analysis, mass variables are used to validate the symmetry with efficiency corrections. The final discriminant uses the score distribution from Neural Network (NN) training as input to the statistical analysis.

Instead of a simultaneous fit for the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ branching ratios, the symmetry method performs independent fits with a single parameter of interest (POI). The results are presented as the branching ratio difference between $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$.

7.3 Cut-based Analysis

The cut-based analysis (CBA) is based on the variable distribution with the best discrimination power. Although this analysis uses the MVA outputs as the final discriminant in the statistical analysis, the variable distributions are used in the validation processes and as the inputs of the MVA.

The reconstructed Higgs boson mass is used in the CBA, the missing mass calculator (MMC) estimator is used in $\ell\tau_{\ell'}$ channel, while the collinear mass approximation is exploited in the $\ell\tau_{had}$ final states. The definition of each mass variable is given in this chapter.

7.3.1 Mass Variables

In the rest frame of the Higgs boson, the τ -lepton and the light lepton are produced back to back, while the decay products of the τ -lepton are boosted along the same direction of the τ . The mass variables can be determined and reconstructed based on this event topology and the invariant mass of the Higgs boson, the mass reconstruction methods considered in this analysis are:

- **Missing Mass Calculator (MMC)**

The MMC is a reconstruction method [138] used in the analysis of the $H \rightarrow \tau\tau$ decay, and it is adapted to the Higgs LFV analysis with less complexity based on one assumption that neutrinos are the only source to count the missing momentum in the single τ decay of the event. This variable has a very good discrimination power in the leplep channel due to the signature of the decay products of τ .

The event topology can be reconstructed by solving a few unknowns such as the x-, y- and z-components of the momentum carried by the neutrinos originate from the τ decays. An under-constrained system has more unknowns than the constraints, the MMC invariant mass is reconstructed by solving an under-constrained system based on the constraints from the

measured x- and y-components of the missing transverse momentum and the visible mass of both the light lepton and τ . A scan is performed over the x- and y-components of the transverse momentum and the yet undetermined variables, based on the event topology of the τ decay and the E_T^{miss} resolution, the weights of each scan point is assigned according to its probability. The MMC invariant mass is estimated by the most probable value of the weighted scan points. The performance of this method is highly relied on the E_T^{miss} resolution. Figure 8.2 (a) shows the MMC mass distribution in the nonVBF region and $e\tau_\mu$ final state.

- **Collinear Mass**

The collinear mass m_{coll} is used in this analysis for validating the symmetry after efficiency corrections (as shown in figure 9.1) and improving the data over background agreements. This mass variable is reconstructed based on the assumptions that the visible decay products of the τ can be used to estimate the τ direction, while the decay products can also be used to derived $\mathbf{p}_T^\tau$ with the E_T^{miss} , in the way that:

$$\begin{aligned} \ell\tau_{\ell'} : \quad \mathbf{p}_T^\tau &= \mathbf{p}_T^{\ell'} + \mathbf{E}_T^{miss} \\ \ell\tau_{had} : \quad \mathbf{p}_T^\tau &= \mathbf{p}_T^{\tau_{had-vis}} + \mathbf{E}_T^{miss} \end{aligned} \quad (7.1)$$

Expressing the four momentum in term of the pseudorapidity and the azimuthal angle, the invariant mass of the system can be obtained as:

$$m_H = \sqrt{2p_T^\ell p_T^\tau (\cos\Delta\eta - \cos\Delta\phi)} \quad (7.2)$$

where $\Delta\eta$ and $\Delta\phi$ are the difference in the pseudorapidity and azimuthal angle between the leading lepton and the sub-leading lepton, respectively. Therefore, the collinear mass can be defined as:

$$\begin{aligned} \ell\tau_{\ell'} : \quad m_{coll} &= \sqrt{2p_T^{\ell_1} (p_T^{\ell_2} + E_T^{miss}) (\cos\Delta\eta - \cos\Delta\phi)} \\ \ell\tau_{had} : \quad m_{coll} &= \sqrt{2p_T^\ell (p_T^{\tau_{had-vis}} + E_T^{miss}) (\cos\Delta\eta - \cos\Delta\phi)} \end{aligned} \quad (7.3)$$

where ℓ_1 and ℓ_2 are the leading and subleading leptons in the leplep channel. Figure 8.3 (b) shows the collinear mass distribution in the $\mu\tau_{had}$ final state and VBF region.

- **Visible Mass**

The visible mass m_{vis} is reconstructed based on the visible decay products. In the leplep channel, it is defined as the invariant mass of the two leptons. In the lephad channel, it is defined as the invariant mass of the light lepton and the $\tau_{had-vis}$. Figure 8.3 (a) shows the visible mass distribution in the $\mu\tau_{had}$ final state and nonVBF region.

- **Transverse Mass**

The transverse mass m_T of a particle is defined by equation 4.5. The transverse mass between a light lepton and the τ is defined as:

$$m_T(\ell, \tau) = \sqrt{2p_T(\ell)p_T(\tau)(1 - \cos\Delta\phi)} \quad (7.4)$$

While the transverse mass between a lepton and the E_T^{miss} is also used in the analysis, and it can be defined as:

$$m_T(\ell, E_T^{miss}) = \sqrt{2p_T(\ell)E_T^{miss}(1 - \cos\Delta\phi)} \quad (7.5)$$

where $\Delta\phi$ is the difference in the azimuthal angle between the lepton $\ell = e, \mu, \tau$ and the E_T^{miss} . This variable for the $e\tau_\mu$ final state and nonVBF region is shown in figure 8.2 (b).

Chapter 8

Event Selection and Categorization

In this chapter, event selection criteria and categorization of the events defined in the MC-template method are introduced, event selections implemented in the symmetry method with the $\ell\tau_{had}$ final states are also discussed. The events from the detector should be selected according to the purposes of reducing the major backgrounds, enhancing the sensitivity and being analysis specific. Events are categorized according to the signature of the Higgs production modes, details are given in section 8.1.6.

8.1 Selection of the $\ell\tau_{\ell'}$ Channel

In the $\ell\tau_{\ell'}$ channels, the selected events must contain one electron and one muon with opposite-sign according to the topology of the Higgs LFV signal, and all events with τ_{had} should be rejected. In most cases, the lepton with higher p_T corresponds to the prompt lepton while the other one with lower p_T originate from the τ -lepton decay. In the following context, the leading lepton is denoted as ℓ_1 and the subleading lepton is denoted as ℓ_2 .

8.1.1 Trigger Selection

Both single-lepton triggers and dilepton triggers are utilized to select $\ell\tau_{\ell'}$ event candidates. Different triggers were chosen for distinct data-taking periods from 2015 to 2018. Since varying leptons are chosen in different analyses, a trigger matching procedure is necessary for the selected electrons and muons in this analysis. Table 8.1 displays the trigger selection in the $\ell\tau_{\ell'}$ channels.

Each trigger is associated with a corresponding p_T threshold, which is applied to the matched offline-reconstructed lepton. Differing p_T thresholds are applied to both leptons for the dilepton triggers. These offline thresholds are also outlined in table 8.1.

The trigger selection is predicated on a logical OR operation among the decisions of the single-electron, single-muon, and electron-muon triggers, alongside the p_T threshold conditions. A decision is considered met when each trigger and its associated threshold are concurrently satisfied. This OR strategy enhances the signal acceptance compared to the strategy employed in the 2015-2016 analysis [49].

The name of each trigger shown in table 8.1 contains the information of the lepton, the p_T threshold and identification working point requirements, a full description of the trigger names are given in [139, 140].

Data period	Trigger type	Chain name	Offline thresholds
2015	Single electron	HLT_e24_lhmedium_L1EM20VH, HLT_e60_lhmedium or HLT_e120_lhloose	$p_T^e > 25$ GeV
	Single muon	HLT_mu20_l1loose_L1MU15 or HLT_mu50	$p_T^\mu > 21$ GeV
	Electron-muon	HLT_e17_lhloose_mu14	$p_T^e > 18$ GeV, $p_T^\mu > 15$ GeV
2016	Single electron	HLT_e26_lhtight_nod0_ivarloose, HLT_e60_lhmedium_nod0 or HLT_e140_lhloose_nod0	$p_T^e > 27$ GeV
	Single muon	HLT_mu26_ivarmedium or HLT_mu50	$p_T^\mu > 27$ GeV
	Electron-muon	HLT_e26_lhmedium_nod0_L1EM22VHI_mu8_noL1	$p_T^e > 27$ GeV, $p_T^\mu > 10$ GeV
2017–2018	Single electron	HLT_e26_lhtight_nod0_ivarloose, HLT_e60_lhmedium_nod0 or HLT_e140_lhloose_nod0	$p_T^e > 27$ GeV
	Single muon	HLT_mu26_ivarmedium or HLT_mu50	$p_T^\mu > 27$ GeV
	Electron-muon	HLT_e26_lhmedium_nod0_L1EM22VHI_mu8_noL1 or HLT_e17_lhloose_nod0_mu14	$p_T^e > 27$ GeV, $p_T^\mu > 10$ GeV $p_T^e > 18$ GeV, $p_T^\mu > 15$ GeV

Table 8.1: Trigger selections used in the $\ell\tau_{\ell'}$ channels with an OR strategy. The associated p_T threshold for each trigger is also shown.

8.1.2 Object Selection

In addition to the event reconstruction and trigger selections, the final state objects are further selected according to some criteria based on analysis. In all object selections, a minimum requirement of two tracks associated with the primary vertex is applied. An event cleaning procedure is applied to remove the problematic events seen in 2015 and 2016 data, a cut is applied to those events with poor quality in the electromagnetic endcap region, only a negligible amount of events were removed. The baseline object selections and the further requirements on the objects surviving the OLR are summarised as follow:

- **Electrons**

Baseline electrons are selected with the criteria of $p_T > 15$ GeV and $|\eta| < 2.47$. The transition region between end-cap and barrel calorimeters within $1.37 < |\eta| < 1.52$ is a crack region with a large amount of inactive materials, and electrons inside the crack region are vetoed.

Baseline electrons are input to the OLR procedure as shown in table 6.2, for those electrons which surviving the OLR are further required to pass the MEDIUM identification and GRADIENT isolation criteria.

For ensuring the trajectories of the electrons are consistent with the tracks originate from the hard-scattering vertex, the longitudinal impact parameter of the electron tracks is required to have $|z_0 \sin\theta| < 0.5$ mm, and the significance of the transverse impact parameter $|\sigma(d_0)/d_0| < 5.0$. However, the electrons from a τ_ℓ decay is expected to be displaced with respect to the primary vertex, which reduces a non-negligible amount of events, these cuts are only applied under certain conditions.

The selected electrons in the symmetry method should satisfy the same criteria above, except that the cut on the significance of the transverse impact parameter is higher $|\sigma(d_0)/d_0| < 10$ in order to allow for the origin from a τ_{had} decay.

- **Muons**

Baseline muons are selected with the criteria of $p_T > 10$ GeV and $|\eta| < 2.47$, they are also input to the OLR procedure. Muons surviving the OLR have to further satisfy the MEDIUM identification and FCTIGHT_FIXEDRAD isolation criteria. The same requirements of the impact parameters applied on the electrons are also applied to the muons under certain conditions.

Selected muons in the symmetry method are also required to satisfy the criteria above, except that the isolation working point is *FCTightTrackOnly*, which tightens the thresholds for the track-based isolation. As the electrons and muons should be selected symmetrically, the muons in the crack region with $1.37 < |\eta| < 1.52$ are also vetoed.

- **Taus**

Baseline $\tau_{had-vis}$ objects are used for E_T^{miss} calculation and the hadronic tau veto in the lelep channels, they are further required to satisfy the MEDIUM tau identification working point and the LOOSE EBDT working point and input to the OLR procedure.

- **Jets**

Baseline jets are selected with the criteria of $p_T > 20$ GeV and $|\eta| < 4.5$, they are further required to satisfy the LOOSEBAD cleaning working point and input to the OLR procedure. JVT algorithm is applied to the jets surviving the OLR to suppress the jets originate from pileup. A b-veto cut is applied to remove the jets initiated by b-quarks. Selected jets in the symmetry method are required to satisfy the criteria above.

8.1.3 Lepton Assignment

The signature of the final state objects in $e\tau_\mu$ and $\mu\tau_e$ channel are the same. In the rest frame of the Higgs boson, light leptons with higher p_T usually correspond to the prompt lepton, and the other originates from the τ decay. But that does not imply that the leading lepton in the rest frame is also the prompt lepton in the laboratory frame. The μe and $e\mu$ channel are defined based on the order of their momentum. The objects from the $e\tau_\mu$ ($\mu\tau_e$) in the μe ($e\mu$) channel can dilute the signal and reduce the sensitivity. The optimization of the separation between the two channels is performed by the lepton assignment procedure, defining the $e\tau_\mu$ and $\mu\tau_e$ channel with the following rules:

- The four-momentum of the E_T^{miss} is built under the assumption that the η value of the neutrinos is the same of the two-leptons system.
- The two leptons and the four-momentum of the E_T^{miss} are used to build the four-momentum of the Higgs boson.
- The invariant mass of the Higgs boson is constrained to $m_H = 125$ GeV.
- In the rest frame of the Higgs boson, the four-momenta of the charged leptons are calculated.
- The lepton with the highest p_T in the rest frame of the Higgs boson is assumed as the prompt lepton.

The μe and $e\mu$ are defined base on the order of momentum, while the $e\tau_\mu$ and $\mu\tau_e$ channel are defined base on the lepton assignment. The p_T ordering in the laboratory frame is significantly improved by the lepton assignment in the rest frame of the Higgs boson, especially for the VBF production mode.

8.1.4 Baseline Selection

Only the events have exactly one reconstructed electron and one muon pass the trigger selection and the object selection criteria given above are considered for the baseline selection. The events

pass the baseline selection are further categorized into VBF and nonVBF. The events only pass the baseline selections are used in the validation regions (VR) but not used in the statistical fit. The events pass the baseline selections and the VBF or nonVBF selections are used in the signal regions (SRs) and used in the statistical analysis. The VBF and nonVBF selections are discussed in section 8.1.6. The baseline selections defined in the lelep channel are summarised in table 8.2, and the details are given as follow:

- Both $e\tau_\mu$ and $\mu\tau_e$ channel are required to have exactly one electron and one muon with opposite sign, and pass the lepton assignment.
- The leading lepton and the subleading lepton are required to have $p_T^{\ell_1} > 45$ GeV and $p_T^{\ell_2} > 15$ GeV, respectively. In symmetry method, the leading lepton is required to have $p_T^{\ell_1} > 35$ GeV.
- In order to suppress the $t\bar{t}$ background, the visible mass of the two-leptons system is required to have $30 < m_{\ell\ell} < 150$ GeV.
- If the subleading lepton is electron, the ratio of the p_T of the tracks to the clusters is required to have $0.2 < \frac{p_T^{track}}{p_T^{cluster}} < 1.25$. This cut is aim to reduce the $Z \rightarrow \mu\mu$ background, one of the muons from this process deposit a large fraction of energy to the calorimeter.

8.1.5 Control Regions

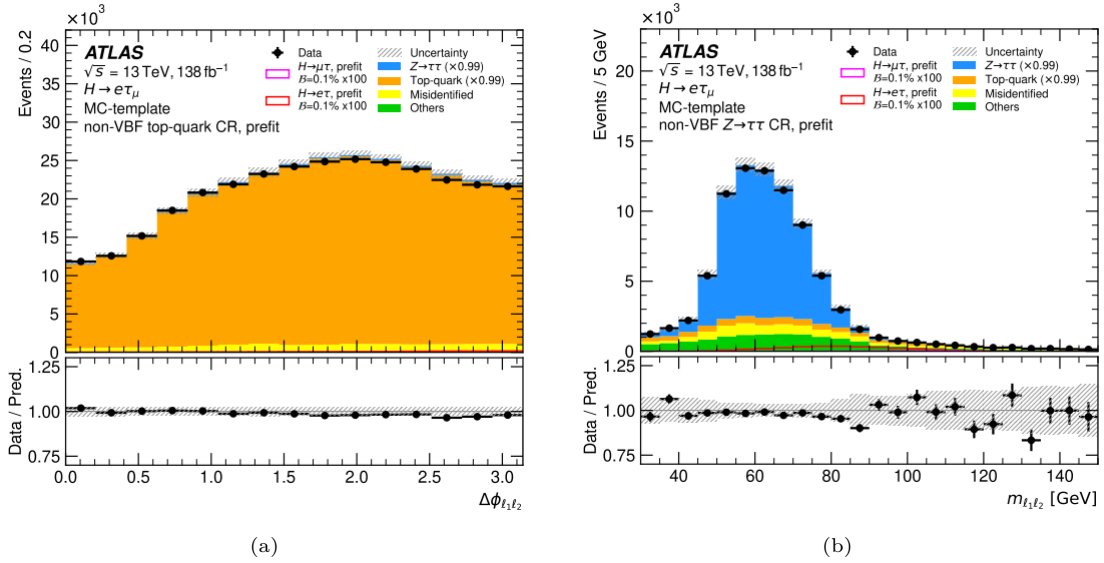


Figure 8.1: The lelep representative distributions for the (a) Top-quark CRs and (b) $Z \rightarrow \tau\tau$ CRs before the statistical analysis, the variables of $\Delta\phi_{\ell_1\ell_2}$ and $m_{\ell_1\ell_2}$ are chosen, respectively. The data over predictions are shown in the pads in the bottom. The prefit statistical, experimental and theoretical uncertainties are shown as hashed bands. The calculated normalization factors are used to scale the $Z \rightarrow \tau\tau$ and Top-quarks contributions [51].

Two main backgrounds in the $\ell\tau_\ell$ channel are the $Z \rightarrow \tau\tau$ process and the Top-quarks process. The modeling between data and MC samples for these two backgrounds should be well-defined. Two control regions (CRs) enriched with the corresponding backgrounds are defined with CR selections.

The CRs are defined per categories per channels, and they should be orthogonal to the signal regions (SRs). Normalization factors are defined in the CRs and used in the statistical analysis.

Events in the Top CRs are required to pass the baseline selections but invert the b-veto criteria by requiring at least one b-tagged jet. This criterion ensures that the purity of the Top CRs is sufficiently high. Figure 8.1 (a) shows the representative distribution of the azimuthal angle between the two leptons, $\Delta\phi_{\ell_1\ell_2}$, for the Top-quark CRs before the statistical analysis. The perfect data-to-prediction ratio is shown in the bottom panel.

Similar to the Top CRs, events in the $Z \rightarrow \tau\tau$ CRs are required to pass the baseline selections, but with a change in the transverse momentum cut to $35 < p_T^{\ell_1} < 45$ GeV. The $Z \rightarrow \tau\tau$ normalization factors are calculated based on the CRs. Figure 8.1 (b) shows the representative distribution of the invariant mass of the two-lepton system, $m_{\ell_1\ell_2}$, for the $Z \rightarrow \tau\tau$ CRs. A good data-to-prediction agreement is achieved by scaling the normalization factors to the respective background contributions.

The diboson processes constitute the relevant background in the signal region, which means that dedicated diboson CRs cannot be defined due to the lack of orthogonality between the CRs and the SRs. Instead, a diboson validation region (VR) is defined to validate the background modeling. Events in the diboson VR should pass the baseline selections and additional criteria are applied to enhance the diboson contribution. Further details are discussed in section 10.1, and figure 10.1 shows the representative distributions for the diboson validation region.

8.1.6 Categorization

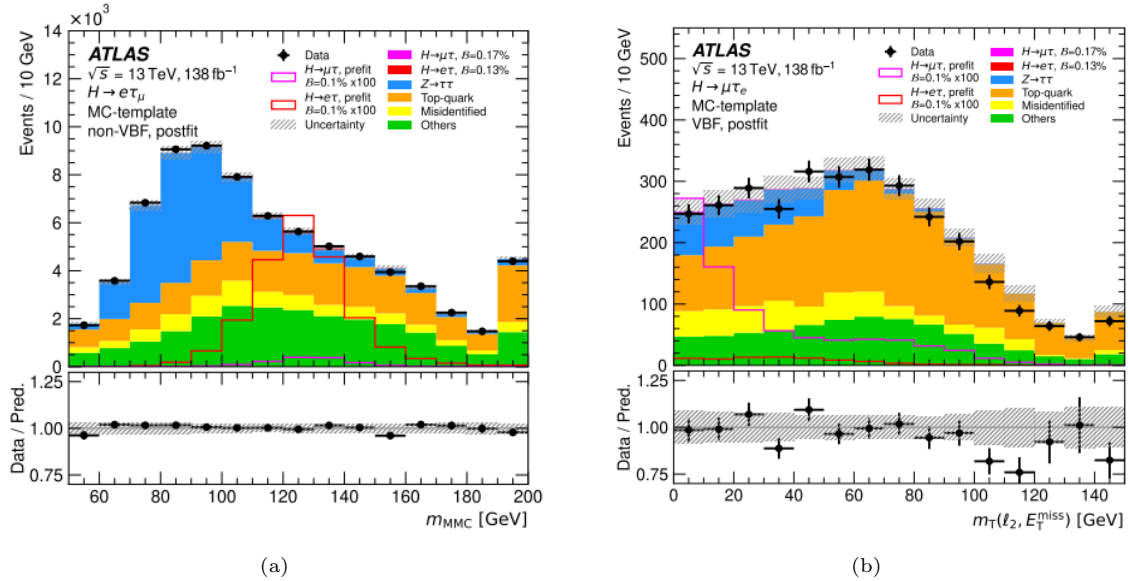


Figure 8.2: The representative kinematic distributions for (a) $e\tau\mu$ and (b) $\mu\tau e$ final states, after a simultaneous fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals. The variable of MMC mass m_{MMC} is chosen for (a) nonVBF, and the transverse mass between the subleading lepton and the E_T^{miss} $m_T(\ell_2, E_T^{miss})$ is chosen for (b) VBF. The hashed band represents the prefit statistical uncertainty, the experimental and theoretical uncertainties. The data over prediction is shown in the bottom pad [51].

Events pass the baseline selections are further classified into two categories: VBF and nonVBF. The VBF region is defined by using selection criteria which enhance the purity of the VBF production mode. As shown in figure 2.2 (b), the clear signature of the two widely separated jets with high

p_T in the VBF production provide a strong event categorization and enhance the signal sensitivity. Events pass the VBF selection are required to satisfy the following criteria:

- Events should pass all the baseline selection criteria.
- At least two jets with $p_T > 30$ GeV are required.
- The transverse momentum of the leading jet should greater than 40 GeV.
- the invariant mass of the dijet system should greater than 400 GeV.
- The absolute value of the pseudorapidity between the two jets should greater than 3.

Events in the nonVBF region should pass the baseline selection but fail the VBF selection. The representative kinematic distributions for the nonVBF and VBF are shown in figure 8.2 (a) and (b), and in the $e\tau\mu$ and $\mu\tau_e$ final states, respectively.

8.2 Selection of the $\ell\tau_{had}$ Channel

In the $\ell\tau_{had}$ channels, the selected events must contain one reconstructed light lepton (electron or muon) and one reconstructed $\tau_{had-vis}$ of opposite sign. Therefore, two channels are considered: the $e\tau_{had}$ and the $\mu\tau_{had}$ channel. In this section, the selections and categorizations in the lephad channel using both MC-template method and symmetry method are introduced. The selection criteria in each case is summarised in Table 8.2.

8.2.1 Trigger Selection

In the lephad channels, the trigger selections are mostly follow the selection used in the lelep channels, as summarised in table 8.1. However, only the single-lepton triggers are considered, no dilepton trigger is used in the lephad channel.

8.2.2 Object Selection

In the lephad channels, at least two tracks associated wit the primary vertex are also required. Event cleaning procedure is applied to the problematic events found in both the 2015 and 2016 data. The object selection of the electrons, the muons and the jets follow the selections used in the lelep channel.

In the lephad channels, only the $\tau_{had-vis}$ candidate with highest transverse momentum is chosen as the baseline $\tau_{had-vis}$ object, while the others are considered as jets, baseline τ leptons are input to the OLR procedure. Those $\tau_{had-vis}$ objects surviving the OLR procedure are further required to satisfy the TIGHT identification working point. The transverse momentum of the selected $\tau_{had-vis}$ objects are additionally required to be greater than 25 GeV in order to suppress the background contributions. Events in the $e\tau_{had}$ channel are further selected by requiring the $\tau_{had-vis}$ objects with one-track to satisfy the MEDIUM working point of the EBDT algorithm.

8.2.3 Baseline Selection

Similar to the lelep channels, only the events have exactly one reconstructed light-lepton and one reconstructed $\tau_{had-vis}$ object passing the trigger selection and the object selection criteria given above are considered for the baseline selection in the lephad channel. The lephad baseline selections are summarised in table 8.2 and the details are given as follow:

- Exactly one reconstructed light-lepton and one reconstructed $\tau_{had-vis}$ of opposite sign.
- The transverse momentum of the lepton is required to have $p_T^\ell > 27.3$ GeV, the transverse momentum and the pseudorapidity of the $\tau_{had-vis}$ objects is required to have $p_T^{\tau_{had-vis}} > 25$ GeV and $|\eta^{\tau_{had-vis}}| < 2.4$, respectively. In symmetry method and for the symmetric reason, the lepton selections are instead required to satisfy $p_T^\ell > 30$ GeV, $|\eta^\ell| < 0.1$ and $|\eta^\ell| < 2.4$, muons in the crack region should also be vetoed $1.37 < |\eta^\ell| < 1.52$.
- In order to suppress the W +jets and the QCD multi-jets background contributions, the sum of the cosine of the azimuthal angle between the light-lepton and the E_T^{miss} is required to have $\Sigma \cos \Delta \phi(\ell, E_T^{miss}) > -0.35$.
- The absolute value of the pseudorapidity between the light-lepton and the $\tau_{had-vis}$ object should be less than 2.
- Only the $\tau_{had-vis}$ objects with one track or three 3 tracks are considered.
- All the b -tagged jets must be rejected in order to suppress the Top processes.
- Electron veto: BDT score > 0.15 for events with 1-prong. In symmetry method, the e-veto cut is also applied to the $\mu\tau_{had}$ channel to obtain the symmetric selection.

8.2.4 Categorization

In the $\ell\tau_{had}$ channels, events are also categorized as VBF and nonVBF. The VBF selection are the same as the selection criteria used in the lelep channels. The two widely separately high-momentum jets also provide a clear signature in the lephad channels, jet-related variables can be used as the inputs of the BDT trainings and significantly increase discrimination power, more details are discussed in chapter 11. Figure 8.3 (a) and (b) show the representative kinematic distributions in the $\mu\tau_{had}$ final state and for the nonVBF and VBF region, respectively.

In the lephad channels, events pass the nonVBF selection criteria should pass the baseline selection criteria and fail the VBF selection. In addition, events are vetoed if the invariant mass between the electron and τ_{had} satisfy $90 < m_{vis}(e, \tau) < 100$ GeV in order to suppress the $Z \rightarrow ee$ processes. In the symmetry method, in order to define a symmetric selection between electrons and muons, additional requirement of $90 < m_{vis}(\mu, \tau) < 100$ GeV is applied.

	$\ell\tau_{\ell'}$		$\ell\tau_{had}$	
	$e\tau_{\mu}$	$\mu\tau_e$	$e\tau_{had}$	$\mu\tau_{had}$
	1e, 1 μ , opposite sign		1 ℓ , 1 $\tau_{had-vis}$, opposite sign	
	Trigger selection			
	$N_{pvx} \geq 1$			
	Batman event cleaning only for 2016 dataset			
	veto $\tau_{had-vis}$			
Baseline Selection	$e\tau_{\mu}$ lepton assignment $p_T^{\ell_1} > 45$ (35) GeV $p_T^{\ell_2} > 15$ GeV $30 < m_{\ell\ell} < 150$ GeV For both ℓ_1 and ℓ_2 : track $ z_0\sin\theta < 0.5$ For $\ell_{prompt} : \text{track } \sigma(d_0)/d_0 < 5$ (10 for both ℓ_1 and ℓ_2) If $\ell_{prompt} = \mu$: electron track $\sigma(d_0)/d_0 < 10$ If $\ell_2 = e : 0.2 < p_T^{track}(\ell_2)/p_T^{cluster}(\ell_2) < 1.25$	$\mu\tau_e$ lepton assignment	1e $p_T^{\ell} > 27.3$ (30) GeV, $(\eta^{\ell} < 2.4)$ $p_T^{\tau_{had-vis}} > 25$ GeV, $ \eta^{\tau_{had-vis}} < 2.4$ $ \Delta\eta(\ell, \tau_{had-vis}) < 2$ $\tau_{had-vis}$ of 1 or 3 tracks $\Sigma\cos\Delta\phi(\ell, E_T^{miss}) > -0.35$ $\tau_{had-vis}$ passes <i>Tight</i> RNN ID	1 μ
	e-veto: BDT score >0.15 for 1-prong (e-veto for $\mu\tau_{had}$)			
	b-veto			
VBF Selection	Pass baseline selection $N_{jets}(p_T > 30) \geq 2$ $p_T^{leadingjet} > 40$ GeV $m_{jj} > 400$ GeV $ \Delta\eta_{jj} > 3$			
nonVBF Selection	Pass Baseline selection, fail VBF selection			
	veto events if			
	$90 < m_{vis}(e, \tau) < 100$ GeV		$(90 < m_{vis}(\mu, \tau) < 100$ GeV)	

Table 8.2: Summary of the event selections of both the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channels, selections in the symmetry method are shown in the brackets.

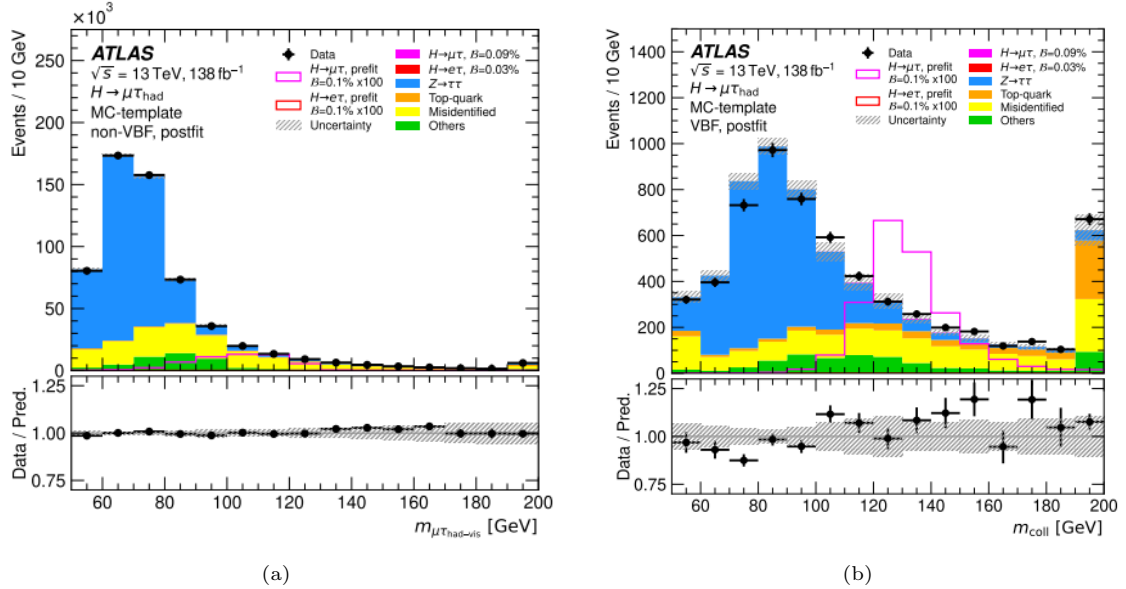


Figure 8.3: The representative kinematic distributions for the $\mu\tau_{had}$ final state, after a simultaneous fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals. The variable of the visible mass m_{vis} is chosen for (a) nonVBF, and the collinear mass m_{coll} is chosen for (b) VBF. The hashed band represents the prefit statistical uncertainty, the experimental and theoretical uncertainties. The data over prediction is shown in the bottom pad [51].

Chapter 9

Efficiency Correction

In the symmetry method, in order to examine the expected symmetry between the SM events which contribute to the $e\tau$ and $\mu\tau$ channels at the interaction points, events are selected with the symmetric criteria discussed in chapter 8. In addition, detector effects invalidate the $e - \mu$ symmetry due to different lepton trigger and lepton detection efficiencies.

The difference in the fake rates could give rise to the unexpected asymmetry between the SM events which contribute to the $e\tau$ and $\mu\tau$ channels, misidentified backgrounds or non-prompt leptons should be subtracted while correcting the difference in efficiency, meaning only the truth events or the *real* SM contributions are considered during the efficiency correction. Different methods were used for the misidentified background estimation as discussed in chapter 10.

In this chapter, the estimation of the muon, electron and trigger efficiency are introduced. The strategies of implementing the efficiency correction to the lelep channels are also discussed. In addition, the current procedure of implementing the efficiency correction to the lephad channels is also described.

9.1 Event Efficiency

The efficiency for the measurements on the events with $\mu\tau_e$ ($\epsilon^{\mu\tau_e}$) or the events with $e\tau_\mu$ ($\epsilon^{e\tau_\mu}$) is the product of the trigger efficiency ϵ_{trig} , the muon ϵ_μ and the electron efficiencies ϵ_e :

$$\epsilon_{tot} = \epsilon_{trig} \cdot \epsilon_e \cdot \epsilon_\mu \quad (9.1)$$

While in the $\ell\tau_{had}$ final states, the total efficiency is defined as:

$$\epsilon_{tot} = \epsilon_{trig} \cdot \epsilon_\ell \quad (9.2)$$

where ℓ represents electron or muon. The considered trigger efficiencies include the single-muon, single-electron and dilepton trigger efficiencies. In the lephad channel, only the single-muon and single-electron trigger efficiencies are considered. As discussed in the chapter 8, the true leptons in the selected events should be reconstructed following the specific criteria used in the analysis. These events should also pass the identification and isolation criteria, the differences of the efficiencies in these processes must be taken into account. Therefore, the muon and electron efficiencies are the product of the efficiencies obtained in the reconstruction, identification and isolation processes:

$$\epsilon_\ell = \epsilon_\ell^{reco} \cdot \epsilon_\ell^{ID} \cdot \epsilon_\ell^{iso} \quad (9.3)$$

9.1.1 Trigger Efficiencies

The trigger selections used in the Higgs LFV analysis are introduced in section 8.1.1. The single-lepton trigger efficiency is defined as the number of triggered lepton candidates divided by the number of isolated lepton candidates, the single-electron and single-muon trigger efficiencies are provided by the ATLAS electron and muon trigger working groups. The ATLAS *TrigGlobalEfficiencyCorrection* tool is used for the global trigger efficiency based on the event topology. The details of the trigger efficiency measurements based on the 2015 dataset are described in [114, 141].

The single-electron trigger efficiencies for the ATLAS Run-2 experiments are described in [142]. The transverse momentum and the pseudorapidity of the electrons are used to parametrize the single-electron trigger efficiencies, in which the efficiencies rise with the p_T from the threshold of the electrons up to nearly 80 GeV, before reaching a stable value of 95% efficiency. The efficiencies are much lower in the crack region $1.37 < |\eta| < 1.52$ and in the region $|\eta| > 2.37$, where a significant amount of inactive materials is present.

The single-muon trigger efficiencies for the ATLAS Run-2 experiments are described in [143]. In the barrel region, the single-muon trigger efficiencies are lower than that of the single-electron trigger, as there is no isolation criteria applied to the offline selected muon in this region. The efficiencies are computed for the barrel and endcap regions separately. The single-muon trigger efficiencies are not parametrized in terms of the transverse momentum of the muons, but in terms of their pseudorapidity and azimuthal angle.

The single-electron and single-muon trigger efficiencies are used to describe the dilepton trigger efficiencies, and the dilepton triggers only used in the lepton channels.

9.1.2 Muon Efficiencies

The muon efficiencies are calculated as the product of the reconstruction, identification and isolation efficiencies of the muons as shown in equation 9.3. The details of the muon efficiency measurements are provided in [120]. The Higgs LFV analysis used the muon efficiencies provided by the ATLAS muon working groups, which were parametrized by the pseudorapidity and azimuthal angle of the muons. Only the muon isolation efficiencies are depending on the transverse momentum of the muons.

The *tag-and-probe* method is used to measure the muon efficiencies in the region with $|\eta| < 2.5$, which is based on the event topology of the $Z \rightarrow \mu\mu$ decays. The first muon candidate in the $Z \rightarrow \mu\mu$ decays is defined as the *tag*, which is required to pass the identification and trigger selection criteria. While the second muon candidate is defined as the *probe* which is used for testing the efficiency of a reconstruction algorithm. Details of this method is given in [120].

9.1.3 Electron Efficiencies

The electron efficiencies are calculated as the product of the reconstruction, identification and isolation efficiencies of the electrons as shown in equation 9.3. The details of the electron efficiency measurements are provided in [114]. The electron efficiencies depend not only the properties of the electrons but also heavily on the kinematic selection. In order to meet the different requirements of

analyses, the ALAS EGamma performance working group did not provide these electron efficiencies in the form of efficiency maps, but provided the scale factors which do not depend on the kinematic selection. The scale factors are calculated for the calibrating the difference in electron efficiencies between data and MC samples.

The Higgs LFV analysis used the electron efficiencies that are calculated based on the provided scale factors (SF) and the MC electron efficiencies from the simulated samples. The MC electron efficiencies are estimated based on the number of electrons in the signal region which pass the reconstruction, identification and isolation criteria, each MC efficiency in the reconstruction algorithm is defined as:

$$\epsilon_i^{MC} = \frac{N_{electrons\ pass\ i}}{N_{all\ electrons}} \quad (9.4)$$

where i represents the reconstruction, identification and isolation. The data electron efficiency is the product of the MC electron efficiency ϵ_i^{MC} and the electron scale factors provided by EGamma working group:

$$\epsilon_i^{Data} = SF_i \cdot \epsilon_i^{MC} \quad (9.5)$$

The selected electron candidates should be the prompt electrons with the matched origin, and the backgrounds that underwent bremsstrahlung conversion and photon conversion should be rejected. The tag-and-probe method used in the muon efficiency measurements is also used for selecting the electrons.

In the leplep channels, the MC electron efficiencies ϵ_i^{MC} are derived by using the TOPQ1 DAODs with the preselection requirements of at least one loose lepton with $P_T > 20$ GeV, and all the SM processes should be available. In the lephad channels, the MC-template method uses the HIGG4D2 DAODs which implemented the *Medium* electron identification working point. However, not all the electrons are stored in the DAODs and the MC electron efficiency cannot be derived. The MC electron efficiencies in the lephad channels should be rederived by using another suitable DAODs in order to retrieve all the electrons and the tau identification information.

9.2 Efficiency Correction Strategies

The assumption of the symmetry between the SM events which contribute to the $e\tau$ and $\mu\tau$ channels at the interaction points imply that the number of SM events N_0 is the same in both channels:

$$N_0 = N_0^{e\tau} = N_0^{\mu\tau} \quad (9.6)$$

where $N_0^{e\tau}$ and $N_0^{\mu\tau}$ represent the number of $e\tau_\mu$ events and the number of $\mu\tau_e$ events in the leplep channels, respectively. That can also refer to the number of $e\tau_{had}$ events and the number of $\mu\tau_{had}$ events in the lephad channels, respectively.

Due to non-perfect detection, the resultant difference in the detection effects give rise to an asymmetry in the number of detected SM events in each channel, which should be corrected via the event efficiency in a way that:

$$N_0 = \frac{N^{e\tau}}{\epsilon^{e\tau}} = \frac{N^{\mu\tau}}{\epsilon^{\mu\tau}} \quad (9.7)$$

Taking the detection effects into account, the number of events in each channel can be retrieved by scaling the event efficiencies, and the expected symmetry can be restored.

The main strategy to retrieve the events is to scale the events one by one, or event-by-event in other word. The efficiencies have dependence on the kinematic properties of the events, the event-by-event correction should be represented as:

$$N_0 = \sum_{i=1}^{N^{e\tau}} \frac{w_i}{\epsilon_i^{e\tau}} = \sum_{i=1}^{N^{\mu\tau}} \frac{w_i}{\epsilon_i^{\mu\tau}} \quad (9.8)$$

where w_i are the weights per event and equal to 1 for the data or the associated MC weight.

Equation 9.7 can be used to obtain the corrected data in both channel. However, in order to perform statistical analysis and the comparison between the uncorrected data and background predictions, the uncorrected $\mu\tau$ data is compared to the efficiency-corrected $e\tau$ data while searching the $H \rightarrow \mu\tau$ signal. In the search of the $H \rightarrow e\tau$ signal, a similar comparison is also performed. According to equation 9.7, the efficiency ratio can be applied in a way that:

$$N^{\mu\tau} = \frac{\epsilon^{\mu\tau}}{\epsilon^{e\tau}} N^{e\tau} = \mathcal{R}^\epsilon \cdot N^{e\tau} \quad (9.9)$$

where $\mathcal{R}^\epsilon$ is the *efficiency ratio* correction factor.

The efficiency correction strategy is proceeded with event-by-event approach, meaning the kinematic properties of the lepton should be included. In leplep channels:

$$\sum_{i=1}^{N^{\mu e}} w_i = \sum_{i=1}^{e\mu} \frac{\tilde{\epsilon}_i^{\mu e}}{\epsilon_i^{e\mu}} w_i = \sum_{i=1}^{N^{e\mu}} \mathcal{R}_i^\epsilon \cdot w_i \quad (9.10)$$

where $\tilde{\epsilon}_i^{\mu e}$ is the relevant μe efficiency associated to the $e\mu$. The μe mirror event to an $e\mu$ event is a virtual event which has the same objects and kinematic properties, except that the roles of the electron and the muon are exchanged. The μe efficiency associated to the real $e\mu$ event can be estimated with the μe mirror event. The efficiency-ratio correction factor applied to the $e\mu$ events is:

$$\mathcal{R}_i^\epsilon = \frac{\tilde{\epsilon}_i^{\mu e}}{\epsilon_i^{e\mu}} = \frac{\epsilon_{trig}^{\mu e}(\{k_i\}_e, \{k_i\}_\mu) \cdot \epsilon_\mu(\{k_i\}_e) \cdot \epsilon_e(\{k_i\}_\mu)}{\epsilon_{trig}^{e\mu}(\{k_i\}_e, \{k_i\}_\mu) \cdot \epsilon_e(\{k_i\}_e) \cdot \epsilon_\mu(\{k_i\}_\mu)} \quad (9.11)$$

where $\{k_i\}_\ell$ represents the kinematic properties of the lepton ℓ .

In the lephad channels, the same strategy used for the leplep channels is used, but the efficiency follows equation 9.2. The event-by-event efficiency-ratio correction is also applied, and the efficiency ratio applied to the $e\tau$ event is:

$$\mathcal{R}^\epsilon = \frac{\epsilon_{trig}^{\mu\tau} \cdot \epsilon_{Reco}^\mu SF_{Reco}^\mu \cdot \epsilon_{ID}^\mu SF_{ID}^\mu \cdot \epsilon_{Iso}^\mu SF_{Iso}^\mu}{\epsilon_{trig}^{e\tau} \cdot \epsilon_{Reco}^e SF_{Reco}^e \cdot \epsilon_{ID}^e SF_{ID}^e \cdot \epsilon_{Iso}^e SF_{Iso}^e} \quad (9.12)$$

where Reco, ID and Iso represents the reconstruction, identification and isolation, respectively.

The efficiency-correction method is validated via the kinematic distributions as shown in figure 9.1, which are the collinear mass distributions created by the simulated events that pass the nonVBF and VBF selection criteria. The validation is currently only done in the leplep channels, and figure 9.1 shows the comparison between the events with a leading muon and a subleading electron from the tau decay $(SM)_{\mu\tau}$, and the events with a leading electron and a subleading muon without

$(SM)_{e\tau}$ or with the efficiency-ratio correction implemented $\mathcal{R}^\epsilon(SM)_{e\tau}$. The solid blue line and the red dots represent the uncorrected SM events of both channels, a clear discrepancy between these two distributions can be found. While the blue dots represented the efficiency-corrected $e\tau$ events, and the symmetry is restored by comparing the red and blue dots which also validated the efficiency-ratio correction factors.

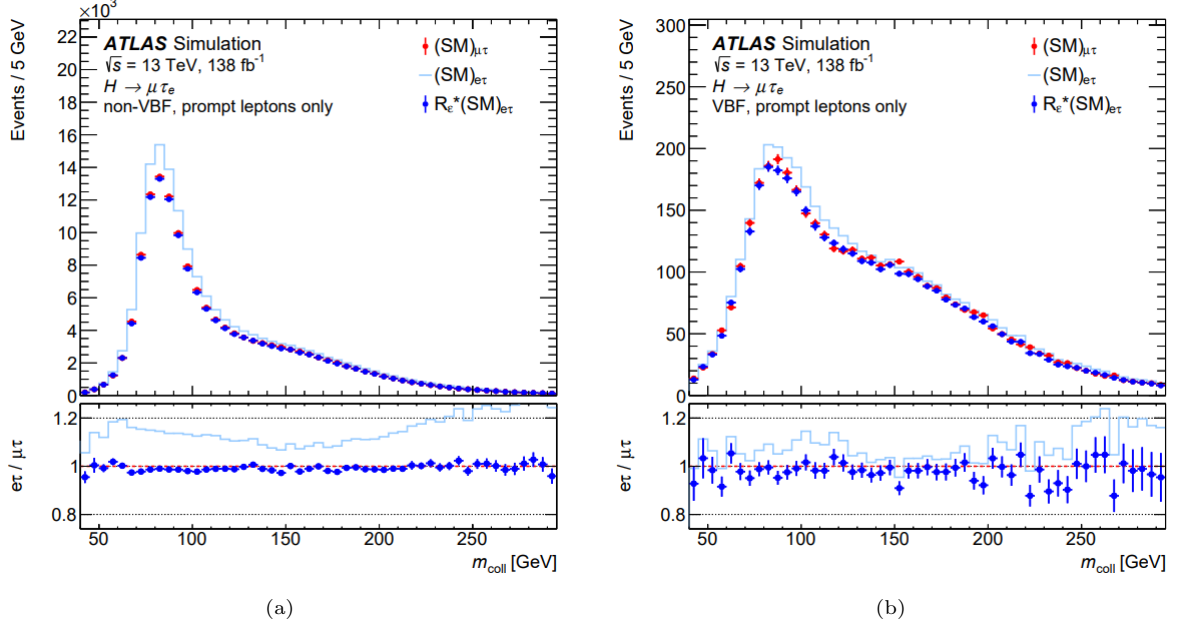


Figure 9.1: The collinear mass distributions in simulated events which pass the (a) nonVBF and (b) VBF selection criteria. The bottom pad shows the ratios of the $e\tau$ over $\mu\tau$, only the statistical uncertainties are included [51].

Chapter 10

Background Estimation

Background processes may have similar signatures that are expected to be observed in the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signal processes. Two background estimation techniques are employed in the Higgs LFV analysis as discussed above: the MC-template method and the symmetry method. In the MC-template method, the data-corrected simulation samples are exploited. In the symmetry method, the backgrounds are estimated via a data-driven approach, simulation samples are used for validations and modelling the signal processes. In both methods and channels, the misidentified backgrounds are estimated via data-driven techniques.

In this chapter, the background estimations via the MC-template method is discussed in section 10.1, and the symmetry method is discussed in section 10.2. Both channels are discussed in each method, however, the studies of the lephad channels in the symmetry method are still in progress at the time of writing, only the general ideas are introduced.

10.1 MC-template Method

The majority of the background contributions in the MC-template method come from the processes of Top-quark, $Z \rightarrow \tau\tau$, $Z \rightarrow \ell\ell$, diboson, SM Higgs and W +jets (only contributed to the lephad channels), which are estimated by MC simulation. Most of the misidentified backgrounds are originated from the QCD and detector processes, and they are difficult to be estimated with high accuracy by simulation. Therefore, data-driven technique is exploited in the fake background estimations.

10.1.1 $\ell\tau_{\ell'}$ Channel

The largest background contributions arise from the processes of Top-quark, $Z \rightarrow \tau\tau$, diboson, and the events with misidentified leptons. Other contributions are the $Z \rightarrow \ell\ell$ processes and the SM Higgs boson decays. As discussed in section 8.1.5, the data-driven control regions and validation regions are defined to model the background processes, and normalization factors are calculated based on the CRs.

Table 10.1 summarises the event yields of the data, background and signal samples in the lelep channels, and for both VBF and nonVBF categories. Large fraction of the events that pass the baseline selection criteria enter the nonVBF region. Each background contribution in the lelep channels are described in order of yields contribution as follow:

- **Top-quark**

The largest background contribution come from the Top-quark processes, which contribute about 32% - 52% of the total background according to table 10.1. Control regions are defined for checking the modelling of the Top-quark background in both the nonVBF and VBF regions. The CRs selection criteria are given in section 8.1.5 and the data over background modelling is shown in figure 8.1 (a).

Normalization factors are calculated in the Top CRs, both the $e\tau_\mu$ and $\mu\tau_e$ channels share the same normalization factor, but different factors are calculated independently for nonVBF and VBF categories due to the difference in the MC normalization. The nonVBF prefit normalization factor is found to be 0.977883, and the VBF prefit normalization factor is 0.941471.

- **Diboson**

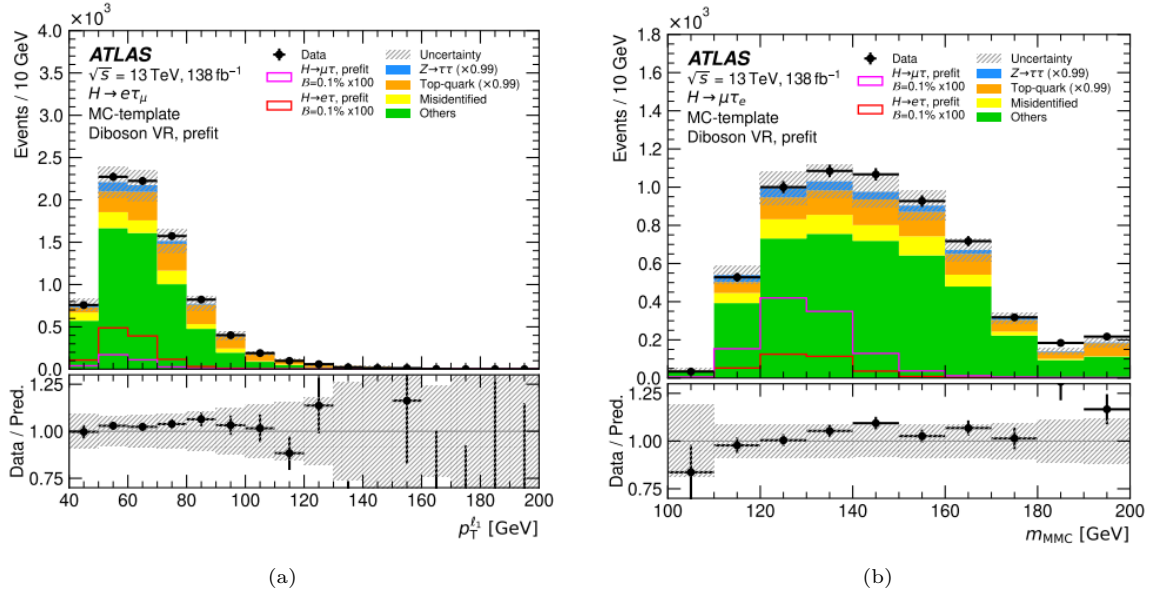


Figure 10.1: Representative distributions for the diboson validation region before statistical analysis: (a) the transverse momentum of the leading lepton in the $e\tau_\mu$ final state, and (b) the MMC mass in the $\mu\tau_e$ final state. The Top-quark and $Z \rightarrow \tau\tau$ contributions are scaled by normalization factors as shown in the legend. The statistical, experimental and theoretical uncertainties are included and indicated by the hashed band [51].

The second large background contribution come from the diboson processes, which has very similar amount of contribution as the Top-quark background in the nonVBF category. As mentioned in section 8.1.5, the diboson processes are estimated by simulation and validated in a dedicated validation region due to the lack of orthogonality to the signal region. In order to ensure the purity in the validation region and reduce other background contribution, the events should pass the baseline selection criteria and the following additional criteria:

- The transverse momentum of the subleading lepton should have $p_T^{\ell_2} > 30$ GeV.
- The visible mass of the two leptons should satisfy $100 < m_{vis} < 150$ GeV.
- The transverse mass between the subleading lepton and the E_T^{miss} : $m_T(\ell_2, E_T^{miss}) > 30$ GeV.

- Jets with $p_T > 30$ GeV are rejected.

The purity of the diboson events in the validation region is around 70%, the good modelling of the diboson validation region are illustrated in figure 10.1, the representative distributions are the transverse momentum of the leading lepton and MMC mass.

- $Z \rightarrow \tau\tau$

The $Z \rightarrow \tau\tau$ processes is also one of the main backgrounds, the modelling of this background should be checked with data. As mentioned in section 8.1.5, dedicated control regions are used to check the modelling and calculate the normalization factors. Figure 8.1 (b) shows the representative distribution of invariant mass of the two-lepton system $m_{\ell_1\ell_2}$ for the $Z \rightarrow \tau\tau$ CRs. The prefit normalization factors are calculated as 0.954092 and 0.678014 for the nonVBF and VBF categories, respectively.

- **Fake lepton background**

In lelep channels, the fake lepton background events consist one or more reconstructed electron or muon originated from photon conversion, heavy-flavour decays, a jet or a τ_{had} misidentified as a prompt light lepton. The main fake lepton candidates come from the W+jets, QCD multi-jets, hadronically decaying $t\bar{t}$ or single Top-quark processes. A data-driven approach which is known as the *ABCD* method is employed to estimate the fake lepton background contribution, the details are discussed in section 10.1.1.1.

- $Z \rightarrow \ell\ell$

The Z boson can decay into two light leptons with the same flavour, a small fraction of the background contributions in the lelep channels come from the $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ processes. In the $Z \rightarrow \mu\mu$ processes, the muons can be misidentified as the electrons by radiating photons while passing through the electromagnetic calorimeter. The $Z \rightarrow \mu\mu$ processes with a muon misidentified as an electron is a sizeable background in the $\mu\tau_e$ channel. The background contribution from the $Z \rightarrow ee$ processes are negligible as the possibility to misidentify an electron as a muon is extremely low.

Dedicated $Z \rightarrow \mu\mu$ control region is defined to study the mismodelling and a normalization uncertainty. To reduce other background contributions, events in the $Z \rightarrow \mu\mu$ control region should pass the baseline selection and following criteria:

- The transverse momentum of the leading lepton should satisfy $35 < p_T^{\ell_1} < 45$ GeV.
- The visible mass of the two leptons should be near the range of Z boson mass peak $75 < m_{vis} < 100$ GeV.
- If the subleading lepton is electron: $1.25 < \frac{p_T^{track}(\ell_2)}{p_T^{cluster}(\ell_2)} < 3$, in order to reject those events without the cluster information and select the events with one of the muons emit photon to misidentify the electrons while passing through the calorimeter.
- The azimuthal angle between the subleading lepton and the E_T^{miss} satisfy $\Delta\phi(\ell_2, E_T^{miss}) < 1.5$.

A good modelling is found in the $Z \rightarrow \mu\mu$ CR but a 25% global normalization offset is observed. The $Z \rightarrow \mu\mu$ CR is not used in the statistical analysis, but a normalization uncertainty is assigned to the $Z \rightarrow \mu\mu$ contribution in the statical analysis.

- **Higgs boson processes**

Only a small fraction of the total background contributions is from the Higgs boson processes, which refers to the SM decays of $H \rightarrow \tau\tau$ and $H \rightarrow WW$. These processes are normalised to their SM predictions.

Sample	$H \rightarrow \mu\tau_e$		$H \rightarrow e\tau_\mu$	
	nonVBF	VBF	nonVBF	VBF
$H \rightarrow \tau\tau$	340 ± 1	27.5 ± 0.3	414 ± 2	32.6 ± 0.2
$H \rightarrow WW$	732 ± 5	72 ± 1	882 ± 3	82 ± 1
$Z \rightarrow \tau\tau + \text{jets}$	17800 ± 100	390 ± 10	22300 ± 73	490 ± 10
$Z \rightarrow \ell\ell + \text{jets}$	1100 ± 100	11 ± 3	700 ± 100	10 ± 3
$t\bar{t}$	29200 ± 40	1700 ± 10	32400 ± 50	1810 ± 10
Diboson	28310 ± 60	630 ± 10	32130 ± 60	670 ± 10
misidentified leptons	12800 ± 200	440 ± 40	10700 ± 300	400 ± 40
Total background	90200 ± 300	3300 ± 40	99500 ± 300	3490 ± 40
$H \rightarrow \mu\tau_e$ signal	1831 ± 9	83 ± 1	144 ± 2	8.7 ± 0.3
$H \rightarrow e\tau_\mu$ signal	138.8 ± 1.0	8.5 ± 0.3	2190 ± 12	95 ± 2
Data	90531 ± 300	3138 ± 60	100769 ± 300	3383 ± 60

Table 10.1: Yields of the $\mu\tau_e$ and $e\tau_\mu$ final states for both the nonVBF and VBF region. $\mathcal{B}(H \rightarrow \mu\tau)$ and $\mathcal{B}(H \rightarrow e\tau)$ are assumed to be 1%. Only the statistical uncertainties are included.

10.1.1.1 Misidentified Lepton Estimation

The background contribution from the events with fake lepton candidates are estimated via a data-driven method. Three dedicated fake control regions are defined for estimating the fake leptons in the signal region, as shown in figure 10.2. In order to establish the fake CRs with enrich fake lepton candidates, the CRs are defined in terms of isolation and identification quality of the subleading lepton, and the y-axis represents the lepton quality criteria. The misidentified lepton is more likely the subleading lepton (96%) than the non-prompt lepton (85%). Therefore, the fake lepton background is studied and estimated in terms of $e\mu$ and μe final states instead of $e\tau_\mu$ and $\mu\tau_e$ channels. In addition, the distribution of the subleading lepton p_T has a large difference between the data in the signal region and the prompt lepton background samples, in which the missing events between them are the fake lepton candidates. The x-axis is defined by the charge composition of the final states, the ratio of the number of events with two light leptons of same-sign charge to the events with opposite-sign charge leptons in the fake CRs is used as a *transfer factor*, under the assumption that the value of this ratio is constant when inverting the isolation quality requirement for subleading lepton.

Each region in figure 10.2 is defined independently as following:

- OSN: opposite sign (OS) nominal region, events should satisfy the same selection criteria as

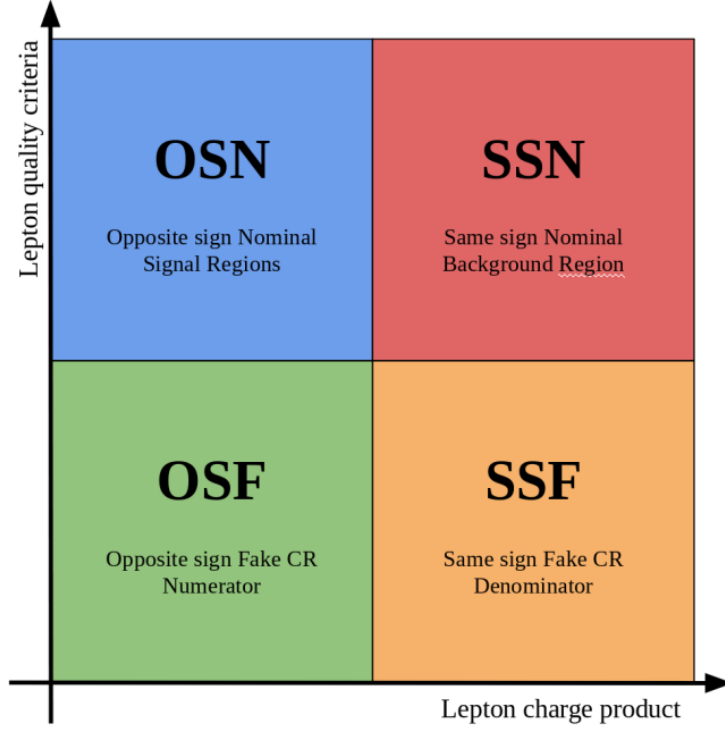


Figure 10.2: A two-dimension view of different regions used in the fake lepton estimation (ABCD method) for the lelep channels.

the SRs, both leptons are isolated.

- SSN: SS nominal region, events should satisfy the same selection criteria as the SRs, except two light leptons with same-sign electric charge is required. Used as CR as it is orthogonal to the signal region.
- OSF: OS fake CR, the numerator of the transfer factor.
- SSF: SS fake CR, the denominator of the transfer factor.

The fake control regions are defined by the isolation and identification criteria of the subleading lepton for increasing the purity of the misidentified leptons. For the events in the fake CRs, the leading lepton should satisfy *Medium* identification and *Gradient* isolation requirements. If the subleading lepton is muon, it should fail the *Gradient* isolation and pass the *Medium* identification requirements. While the subleading lepton is electron, it should fail the *Gradient* isolation or *Medium* identification but still pass the *Loose* identification requirements.

The transfer factors are defined independently for both channels, the number of events with a misidentified lepton in the signal regions and the transfer factors are defined as:

$$N_{fake\ lepton}^{SRs} = N_{Nom}^{SS} \times f_{trans}, \quad f_{trans} = \frac{N_{fakeCR}^{OS}}{N_{fakeCR}^{SS}} \quad (10.1)$$

where $N_{fake\ lepton}^{SRs}$ is the number of events with fake lepton in the SRs, N_{Nom}^{SS} is the number of events in the same-sign nominal region and f_{trans} is the transfer factor.

10.1.2 $\ell\tau_{had}$ Channel

In the lephad channels, the largest background contributions come from the $Z \rightarrow \tau\tau$ processes and events with misidentified τ_{had} . Other contributions are the processes of $Z \rightarrow \ell\ell$, diboson, Top-quark, SM Higgs and W +jets. Dedicated $Z \rightarrow \tau\tau$ normalization factors are calculated at the prefit level.

Table 10.2 summarises the event yields of the data, background and signal samples in the lephad channels, and for both the nonVBF and VBF categories. Similar to lelep channels, large fraction of the events which pass the baseline selection criteria enter the nonVBF region. Each background contribution in the lephad channels are described as follow:

- $Z \rightarrow \tau\tau$

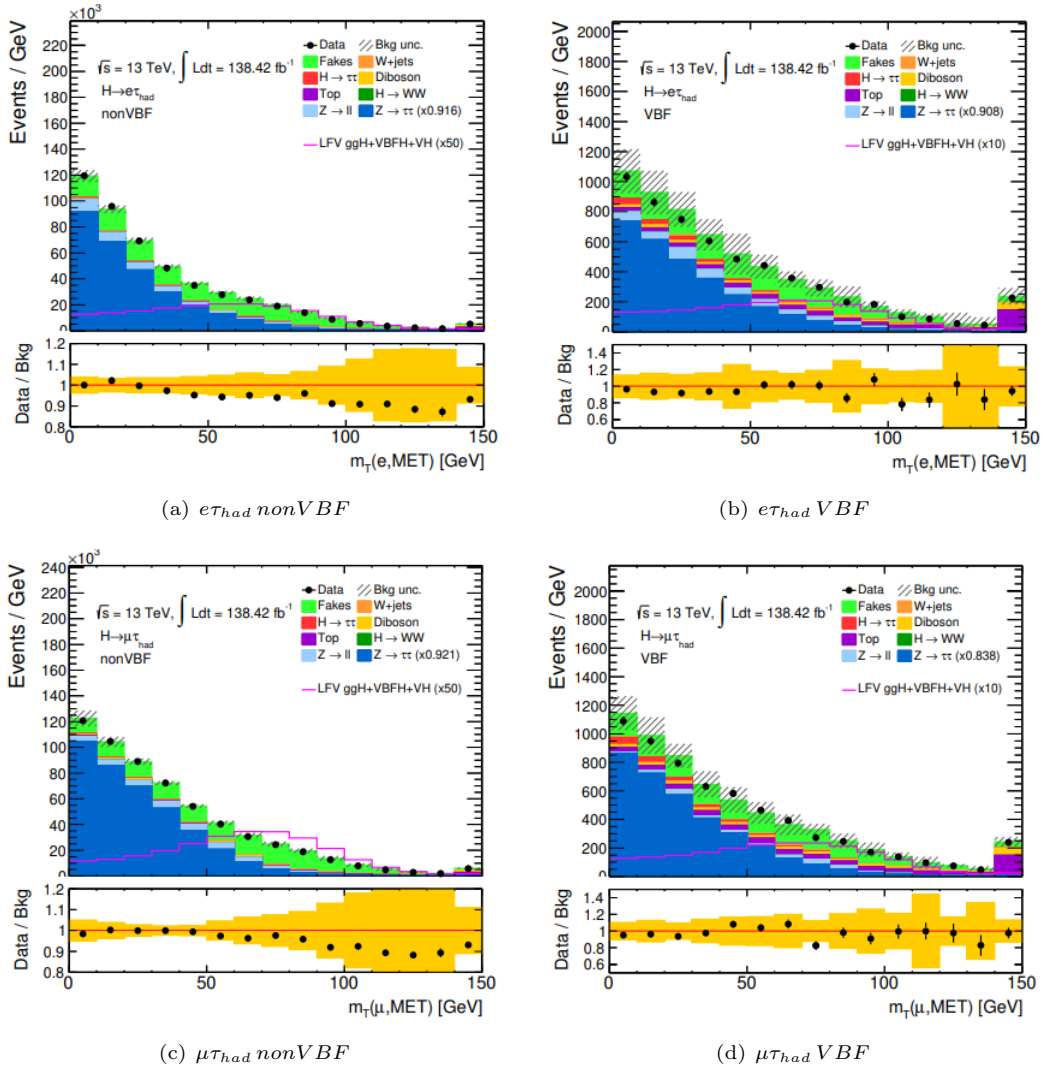


Figure 10.3: The prefit distribution of the transverse mass between the light lepton and E_T^{miss} in (a) $e\tau_{had}$ nonVBF category, (b) $e\tau_{had}$ VBF category, (c) $\mu\tau_{had}$ nonVBF category and (d) $\mu\tau_{had}$ VBF category. Statistical uncertainties plus a subset of the systematic uncertainties are included. The $Z \rightarrow \tau\tau$ samples are scaled by the corresponding NFs as shown in the legends.

The largest background contribution is from the $Z \rightarrow \tau\tau$ processes, which contribute about

48% - 67% of the total background, depending on the region and channel. In the lephad channel, there is no dedicated control region defined to constrain the normalization, but $Z \rightarrow \tau\tau$ normalization factors were calculated based on the BDT score distribution at the prefit (before statistical analysis) level.

As shown in figure 11.5, each event is assigned with a BDT score within the range from -1 to 1, and there are 20 bins in total. The first bin of the BDT score distributions in each channel and category are used to calculate the corresponding $Z \rightarrow \tau\tau$ normalization factors. The data and background events in the first bin are used to calculate the factors as: $NF_{Z \rightarrow \tau\tau} = \frac{N_{Data} - N_{other\ bkg}}{N_{Z \rightarrow \tau\tau}}$, where $N_{other\ bkg}$ refers to the background processes other than $Z \rightarrow \tau\tau$.

The accuracy of the normalization factors are guaranteed by the enrich $Z \rightarrow \tau\tau$ contribution in the first bin of each BDT score distribution, good modelling are achieved as shown in figure 10.3. The normalization factors are calculated independently in the $e\tau_{had}$ nonVBF, $e\tau_{had}$ VBF, $\mu\tau_{had}$ nonVBF and $\mu\tau_{had}$ VBF as: 0.916, 0.908, 0.921 and 0.838, respectively.

- **Misidentified $\tau_{had-vis}$**

The second largest background in the lephad channels are those jets misidentified as hadronically decaying $\tau_{had-vis}$ leptons. The fake $\tau_{had-vis}$ events consist one jet originated from the W +jets or QCD multi-jets processes, and a small fraction of contributions could also be coming from the Top-quark processes. In the W +jets processes, the W boson could decay into one lepton and one neutrino, and the jets could be misidentified as τ_{had} . A data-driven method which is known as the *fake-factor* method is used to estimate the fake $\tau_{had-vis}$ background contribution, the details are discussed in section 10.1.2.1.

- $Z \rightarrow \ell\ell$

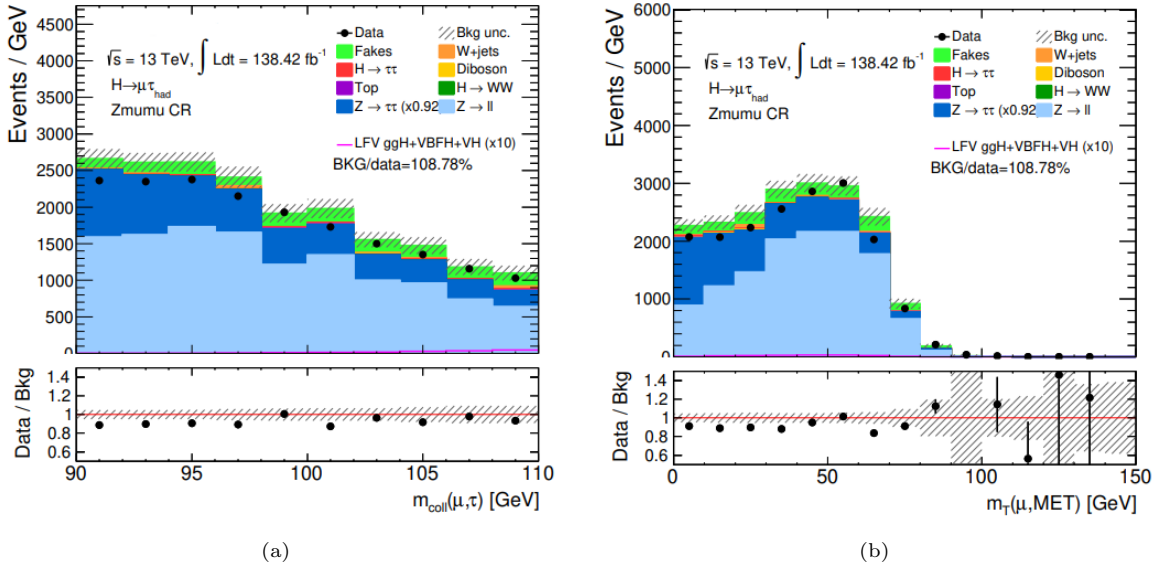


Figure 10.4: The representative distributions of the (a) collinear mass between the muon and $\tau_{had-vis}$, and (b) transverse mass between the muon and E_T^{miss} in the $Z \rightarrow \mu\mu$ validation region. The $Z \rightarrow \tau\tau$ samples are scaled by the NFs. Only statistical uncertainties are shown.

The background contribution in the lephad channels could come from both the $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ processes, and the $Z \rightarrow \mu\mu$ contribute about 3% - 6% of the total background, depending on the category. A $Z \rightarrow \ell\ell$ validation region is defined for validating the modelling of this background, and about 65% of the total contribution is from the $Z \rightarrow \mu\mu$ processes. Events in this validation region should pass the baseline selection and satisfy the following criteria:

- The pseudorapidity of the selected τ leptons should be less than 0.1, $|\eta(\tau)| < 0.1$.
- The collinear mass between the muon and the tau should be near the Z boson mass peak, $90 < m_{coll}(\mu, \tau) < 110$ GeV.

There is no mismodelling observed in this validation region, but a 13% global normalization offset is found. Therefore, an uncertainty of that size is assigned. Figure 10.4 shows the representative distributions for the $\mu\tau_{had}$ final state for events passing the VR, which have good modelling in overall.

- **Diboson**

The diboson processes contribute about 2% - 5% of the total background, depending on the category, which are normalized to their SM predictions.

- **Top-quark**

A small background contribution come from the Top-quark processes in the lephad channels. Top control region is not defined for the lephad channels. However, the normalization factors calculated in the lelep channels are shared with the lephad channels, and they are used in the simultaneous fit on the lelep and lephad channels.

- **W +jets**

The W boson is produced associated with a quark or a gluon, while W boson decays into a neutrino and a light lepton. The quark or gluon forms a jet which could be mimicking the $\tau_{had-vis}$. The W +jets processes give a small background contribution to the lephad channels.

- **Higgs boson processes**

The Higgs boson processes contribute a small fraction of the total background in the lephad channels, the processes are referring to the $H \rightarrow \tau\tau$ and $H \rightarrow WW$. The processes are normalized to their SM predictions.

10.1.2.1 Fake Factor Method

In lephad channels, a data-driven method known as the fake-factor method is employed to estimate the background contribution from the jets misidentified as $\tau_{had-vis}$. In order to estimate the fake tau leptons, dedicated control regions are defined. The $\tau_{had-vis}$ candidates in the control regions should pass all the analysis requirements but fail the *Tight* identification requirement. Therefore, those CRs are also known as the *anti- $\tau_{had-vis}$* (or anti-ID) regions. The misidentified $\tau_{had-vis}$ could be originated from the jets which are induced from quarks or gluons, and candidates with very low identification score are dominated by the gluon-induced and pileup jets. Hence, in the anti- $\tau_{had-vis}$ regions, candidates with tau identification score lower than 0.05 are also accepted. On the other hand, jets in the signal regions are mainly induced from quarks.

Sample	$H \rightarrow \mu\tau_{had}$		$H \rightarrow e\tau_{had}$	
	nonVBF	VBF	nonVBF	VBF
$Z \rightarrow \tau\tau$	423951 ± 300.2	4075.1 ± 28.4	287558.6 ± 253.0	2882.4 ± 21.0
misidentified τ_{had}	144781.4 ± 226.7	1562.7 ± 18	142980.0 ± 234.7	1692.8 ± 18.9
$Z \rightarrow \ell\ell + \text{jets}$	34171.7 ± 629.9	200.7 ± 30.9	34650.8 ± 737.6	430.5 ± 50.1
Diboson	9258.5 ± 29.6	266.2 ± 3.2	7064.6 ± 26.2	288.6 ± 3.3
Top	9054.1 ± 35.5	600.7 ± 9.2	7156.6 ± 32.2	535.4 ± 8.8
$W + \text{jets}$	4262.0 ± 320.4	58.8 ± 11.7	2424.9 ± 240.0	66.7 ± 15.8
$H \rightarrow \tau\tau$	3588.3 ± 8.1	188.9 ± 1.1	2386.6 ± 6.8	166.1 ± 1.1
$H \rightarrow WW$	26.0 ± 0.5	3.9 ± 0.1	3.4 ± 0.2	0.8 ± 0.1
Total background	629093.1 ± 802.0	6957.0 ± 48.1	484225.5 ± 850	6003.2 ± 60.4
LFV ggH	4853.0 ± 6.5	72.3 ± 0.7	3354.0 ± 5.5	65.3 ± 0.7
LFV VBFH	240.3 ± 0.5	137.0 ± 0.4	166.6 ± 0.4	122.0 ± 0.4
LFV VH	150.4 ± 0.7	2.7 ± 0.1	105.8 ± 0.6	2.2 ± 0.1
Data	589198 ± 767.6	6178 ± 78.6	477927 ± 691.3	5713 ± 75.6

Table 10.2: Yields of the $\mu\tau_{had}$ and $e\tau_{had}$ final states for both the nonVBF and VBF region. $\mathcal{B}(H \rightarrow \mu\tau)$ and $\mathcal{B}(H \rightarrow e\tau)$ are assumed to be 1%. Only the statistical uncertainties are included.

The baseline selection criteria of lephad channels required exactly one $\tau_{had-vis}$ for each event, a corresponding anti- $\tau_{had-vis}$ region can be defined for each signal region, which have enrich events with a jet faking the $\tau_{had-vis}$. The anti- $\tau_{had-vis}$ regions are used to estimate the number of events with a jet misidentified as $\tau_{had-vis}$ in the signal region, by defining a transfer factor to correct the difference in the selection efficiency between the pass- $\tau_{had-vis}$ region and the anti- $\tau_{had-vis}$ region, under the assumption that it is a constant value when inverting the tau identification requirement for the $\tau_{had-vis}$.

The transfer factor is also known as the *combined fake-factor* $\mathcal{F}$. In order to estimate the number of events with fake $\tau_{had-vis}$ in the signal region, the number of events with a misidentified $\tau_{had-vis}$ in the anti- $\tau_{had-vis}$ regions are scaled by the combined fake-factor. While the data events with a misidentified $\tau_{had-vis}$ and in the anti- $\tau_{had-vis}$ regions are estimated by subtracting the simulated events without the jets faking $\tau_{had-vis}$ in the anti- $\tau_{had-vis}$ regions:

$$N_{fakes}^{SR} = (N_{Data}^{anti-\tau} - N_{MC, no\ jet \rightarrow \tau}^{anti-\tau}) \times \mathcal{F} \quad (10.2)$$

The combined fake-factor for each signal region is binned in p_T of the $\tau_{had-vis}$ and the number of $\tau_{had-vis}$ tracks.

The jets misidentified as the $\tau_{had-vis}$ could be induced from different sources, as mentioned before, the relevant processes in the lephad channels are the $W + \text{jets}$, QCD multi-jets and Top-quark. Therefore, the fake-factors depend on the quark or gluon composition of a given sample. The fake-factors of each process should be weighted by the corresponding fractional contribution in the

anti- $\tau_{had-vis}$ region:

$$\mathcal{F} = R_W F_W + R_{Top} F_{Top} + R_{QCD} F_{QCD} \quad (10.3)$$

where R is a ratio representing the corresponding fractional contribution, and F is the fake-factor of the respective process. The Top contribution (R_{Top}) is very small in the lephad channels, and the fake-factor of the Top (F_{Top}) production processes is expected to be similar to the F_W . Therefore, most of the misidentified $\tau_{had-vis}$ are quark-initiated jets, and the combined fake-factor can be written as:

$$\mathcal{F} = R_W F_W + R_{QCD} F_{QCD} \quad (10.4)$$

The fake $\tau_{had-vis}$ contributions from the W +jets and QCD multi-jets processes are calculated individually, the sum of the individual ratios should equal to 1: $R_W + R_{QCD} = 1$. The individual jets-faking- $\tau_{had-vis}$ contributions from different processes can be estimated as:

$$N_{fakes_i}^{SR} = (N_{Data}^{anti-\tau} - N_{MC,no\,jet\rightarrow\tau}^{anti-\tau}) \times R_i \times F_i \quad (10.5)$$

where $i = W$ or QCD. To obtain the fake contributions from each process in the signal region, data events from each process with jets-faking- $\tau_{had-vis}$ in the anti- $\tau_{had-vis}$ regions are obtained by scaling the data events with the fractional contribution R_i , and then multiply by the transfer factor F_i . In addition, dedicated W +jets and QCD control regions are defined for estimating the individual fake-factors. Events in each control region should pass the baseline selection criteria, additional criteria are implied for enhancing the relevant backgrounds. As mentioned in section 8.2.3, the cut $\Sigma \cos \Delta \phi(\ell, E_T^{miss}) > -0.35$ is implied to suppress the W +jets background contributions, in the W +jets CRs, this cut is inverted to enhance the W +jets background. In addition, events in the W +jets CRs are further required to satisfy $m_T(\tau, E_T^{miss}) > 40$ GeV. While in the QCD CRs, another baseline selection criteria is inverted for enhancing the QCD multi-jets background: $|\Delta \eta(\tau, \ell)| \geq 2$. To ensure the orthogonality between the two CRs, events in the W (QCD) CRs should satisfy the criteria $m_T(\ell, E_T^{miss}) > 60$ GeV (< 60 GeV). This cut is optimized for obtaining the highest purity of each CR, the purities are calculated in each final state with 1-prong or 3-prong and in each category, for example, the purity of both the W +jets and QCD CRs and pass the anti- $\tau_{had-vis}$ selection is 97% for the $\ell\tau_{1-prong}$ final states. The W +jets and QCD CRs are summarised in table 10.3.

Control Region	Definition with respect to SRs
W +jets CR	$\Sigma \cos \Delta \phi(\ell, E_T^{miss}) \leq -0.35$, $m_T(\tau, E_T^{miss}) > 40$ GeV, $m_T(\ell, E_T^{miss}) > 60$ GeV
QCD CR	$ \Delta \eta(\tau, \ell) \geq 2$, $m_T(\ell, E_T^{miss}) < 60$ GeV

Table 10.3: Selection criteria of the W +jets and QCD control regions.

The individual transfer factors or the fake-factors of each process are defined as follow:

$$F_i = \frac{N_{Data}^{\tau-ID,CR_i} - N_{MC,no\,jets\rightarrow\tau}^{\tau-ID,CR_i}}{N_{Data}^{anti-\tau,CR_i} - N_{MC,no\,jets\rightarrow\tau}^{anti-\tau,CR_i}} \quad (10.6)$$

where $i = W$ or QCD, and $\tau - ID$ represents the events pass the τ *Tight* identification requirement. Figure 10.5 shows the values of F_W and F_{QCD} for the nonVBF and VBF categories, in the $e\tau_{had}$ and $\mu\tau_{had}$ channels, as a function of $\tau_{had-vis}$ p_T for the $\tau_{1-prong}$ and $\tau_{3-prong}$ final states. The binning of the fake-factors was optimized in each category and channel separately, according to the available statistics.

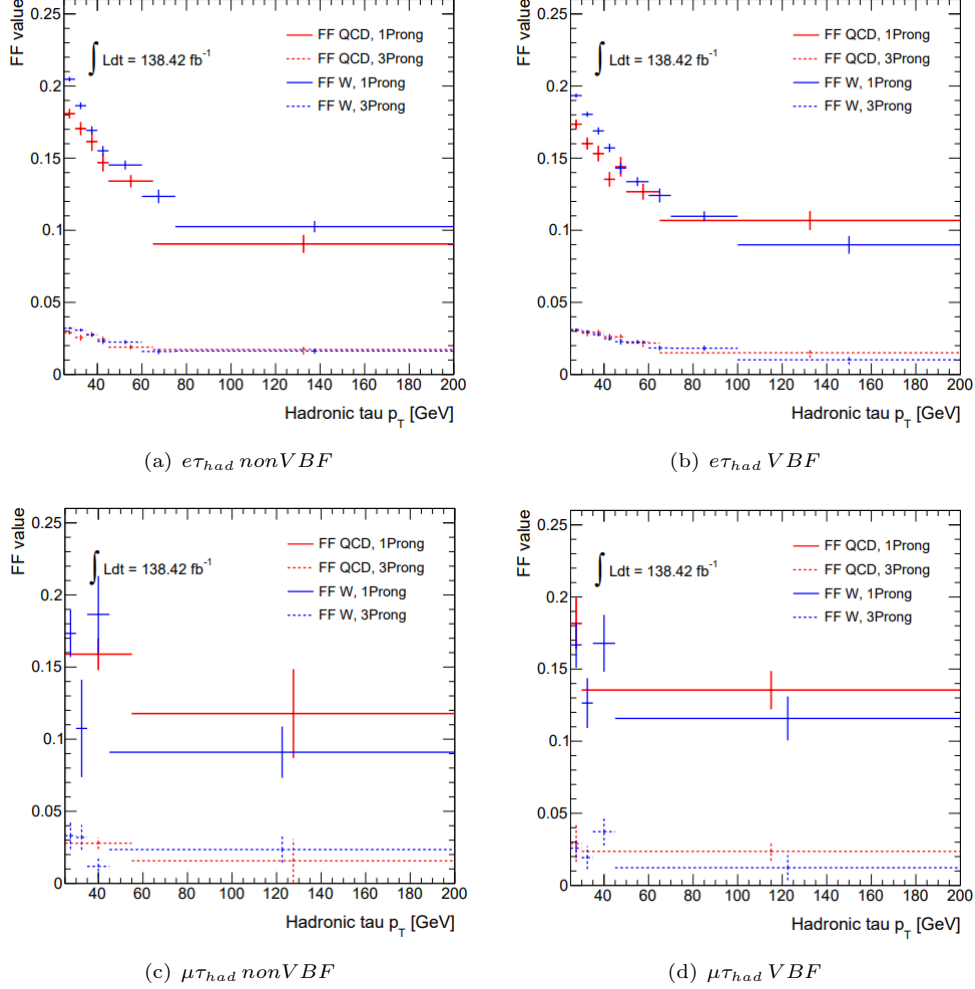


Figure 10.5: F_W and F_{QCD} as a function of $\tau_{had-vis}$ p_T , for $\tau_{1-prong}$ and $\tau_{3-prong}$ final states, and for the (a) nonVBF, $e\tau_{had}$ channel, (b) VBF, $e\tau_{had}$ channel, (c) nonVBF, $\mu\tau_{had}$ channel, (d) VBF, $\mu\tau_{had}$ channel.

The fraction of the QCD multi-jets contribution R_{QCD} in the anti- $\tau_{had-vis}$ regions is estimated by data, the ratio is defined as the number of QCD data events to the total number of data events with jets misidentified as $\tau_{had-vis}$ in the anti- $\tau_{had-vis}$ regions:

$$R_{QCD} = \frac{N_{QCD,Data}^{anti-\tau}}{N_{Data}^{anti-\tau} - N_{MC,no jets \rightarrow \tau}^{anti-\tau}} \quad (10.7)$$

While the value of R_W is obtained by $1 - R_{QCD}$. The number of data events in the QCD CRs fail the τ identification requirement $N_{Data}^{QCD CR, anti-\tau}$ is not necessary the same as the number of QCD data events in the anti- $\tau_{had-vis}$ regions $N_{QCD,Data}^{anti-\tau}$, as events with a true lepton may also have contribution. A transfer factor, called the *Isolation Factor* $\mathcal{I}$, is used to estimate the $N_{QCD,Data}^{anti-\tau}$:

$$N_{QCD,Data}^{anti-\tau} = (N_{Data}^{QCD CR,anti-\tau} - N_{MC,true lepton}^{QCD CR,anti-\tau}) \times \mathcal{I} \quad (10.8)$$

The isolation factor $\mathcal{I}$ is the ratio of the data events passing the lepton isolation requirement over those fail it in a dedicated control region. Events in this CR pass the baseline selection criteria but inverted the opposite-sign requirement to same-sign. This CR is known as the *SSB* region, which is orthogonal to the region where the isolation factor is applied. This transfer factor is also defined under an assumption that its values have no significant difference between the baseline region and the signal regions where they are applied. The data events without true lepton in the SSB region are obtained by subtracting the simulated events with a true lepton, the isolation factor is defined as:

$$\mathcal{I} = \frac{N_{Data}^{iso,SSB} - N_{MC,true lepton}^{iso,SSB}}{N_{Data}^{anti-iso,SSB} - N_{MC,true lepton}^{anti-iso,SSB}} \quad (10.9)$$

where N^{iso} and $N^{anti-iso}$ represents the events pass and fail the lepton isolation requirement, respectively.

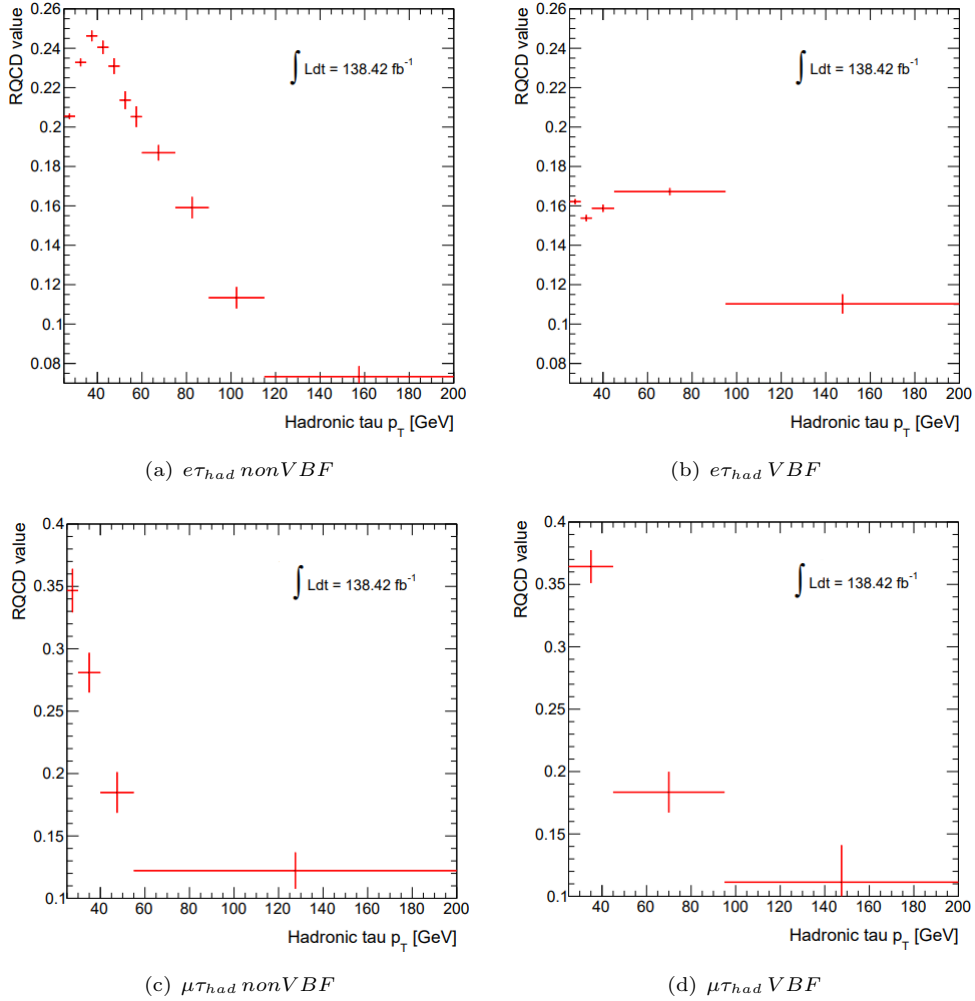
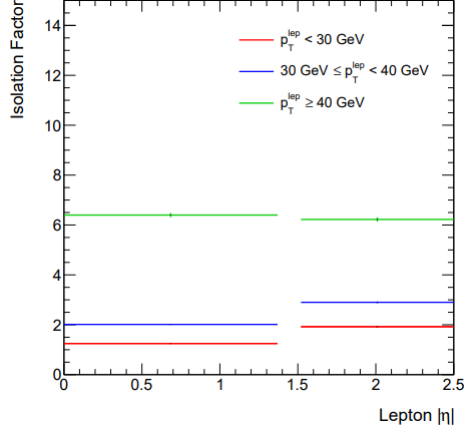
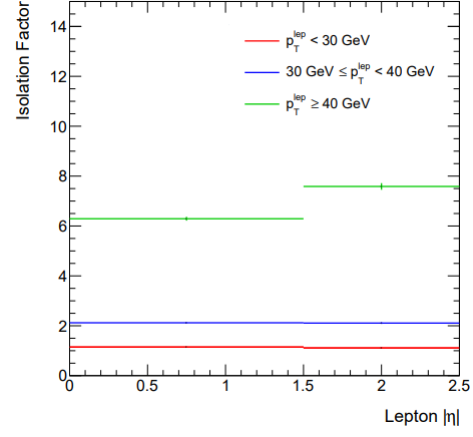


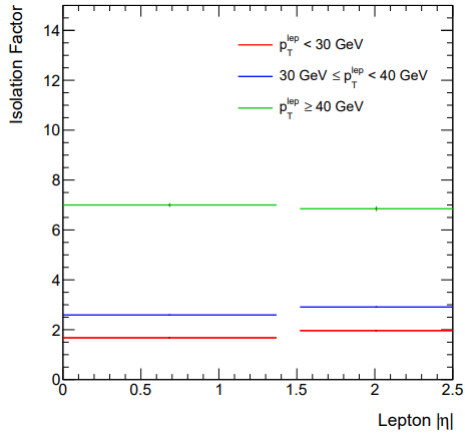
Figure 10.6: The values of RQCD as a function of $P_T^{\tau_{had-vis}}$, for the (a) nonVBF, $e\tau_{had}$ channel, (b) VBF, $e\tau_{had}$ channel, (c) nonVBF, $\mu\tau_{had}$ channel, (d) VBF, $\mu\tau_{had}$ channel.



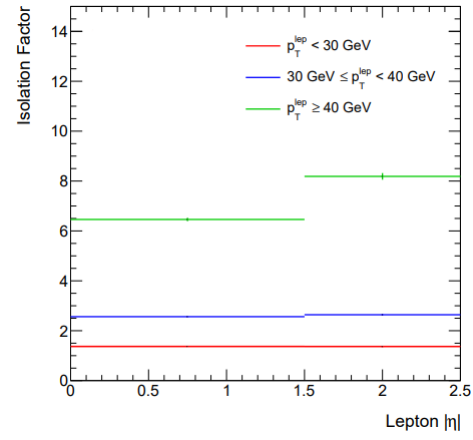
(a) $e\tau_{had}$, 2015 – 2016 data



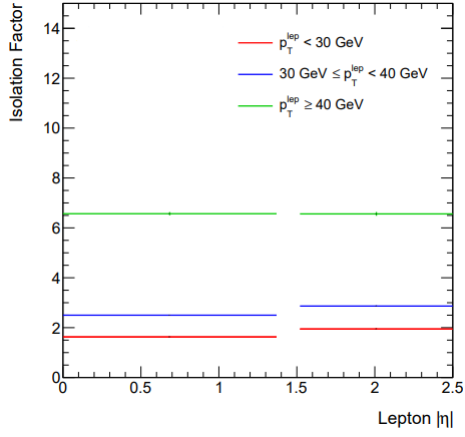
(b) $\mu\tau_{had}$, 2015 – 2016 data



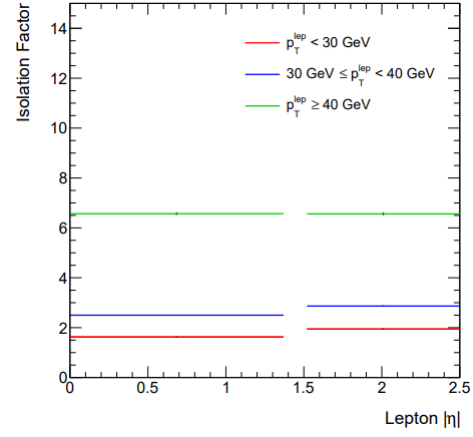
(c) $e\tau_{had}$, 2017 data



(d) $\mu\tau_{had}$, 2017 data



(e) $e\tau_{had}$, 2018 data



(f) $\mu\tau_{had}$, 2018 data

Figure 10.7: Isolation factor as a function of the p_T^ℓ and $|\eta^\ell|$, and of data campaign, for both the $e\tau_{had}$ and $\mu\tau_{had}$ channels, in the baseline region.

Figure 10.6 shows the values of the R_{QCD} in each channel and category, as a function of $\tau_{had-vis}$ p_T . While in the $H \rightarrow \tau\tau$ analyses, the value of R_{QCD} is estimated as function of both $|\Delta\phi(\tau, E_T^{miss})|$ and $\tau_{had-vis}$ p_T . However, a low dependence of the $|\Delta\phi(\tau, E_T^{miss})|$ variable and limited statistics are found in the lephad channels of the Higgs LFV analysis, the binning was optimized according to available statistics instead.

The values of the isolation factors are shown in figure 10.7, as a function of p_T^ℓ and $|\eta^\ell|$, and for both $e\tau_{had}$ and $\mu\tau_{had}$ channels. The isolation factors are calculated separately for each data campaign as incompatibility was found between each dataset from 2015-2018, due to a slight difference in the isolation requirement of the triggers.

The modelling of the W +jets CRs are checked by estimating the fake backgrounds in CRs under the assumption that all contribution is originated from the W +jets processes, therefore, R_W is set to one and R_{QCD} is set to zero. Figure 10.8 shows the $m_{coll}(\ell, \tau_{had-vis})$ distributions for each channel and category, a nearly perfect modelling is achieved in all the distributions in the W +jets CRs.

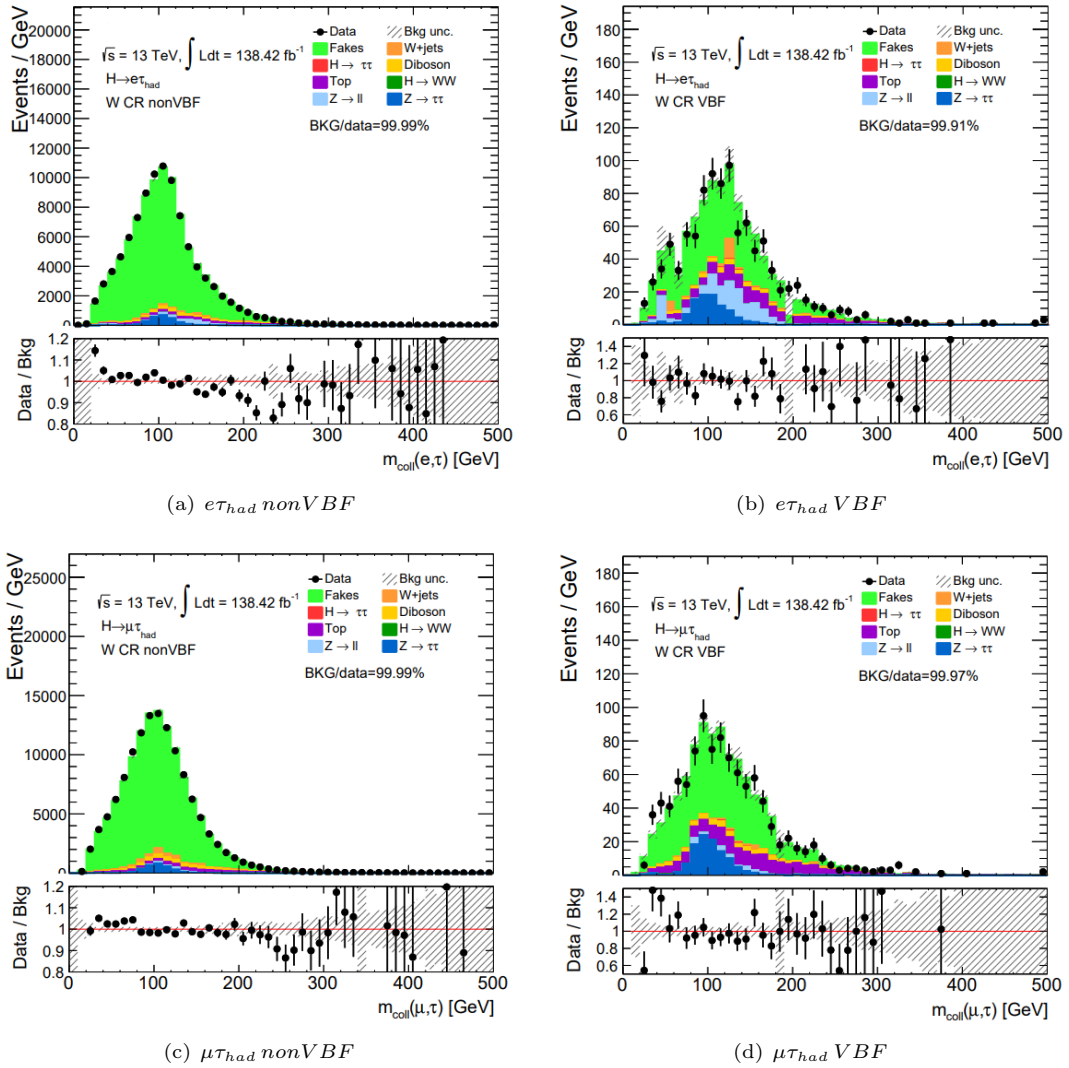


Figure 10.8: The $m_{coll}(\ell, \tau_{had-vis})$ distributions in the W +jets control regions, for the (a) nonVBF, $e\tau_{had}$ channel, (b) VBF, $e\tau_{had}$ channel, (c) nonVBF, $\mu\tau_{had}$ channel, (d) VBF, $\mu\tau_{had}$ channel.

The modelling of the QCD CRs are also checked by estimating the fake backgrounds in the CRs under the assumption that all contribution is originated from the QCD multi-jets processes, and similarly, the R_{QCD} is set to one and R_W is set to zero. Perfect modelling can also be found in the $m_{coll}(\ell, \tau_{had-vis})$ distributions in the QCD CRs, as shown in figure 10.9.

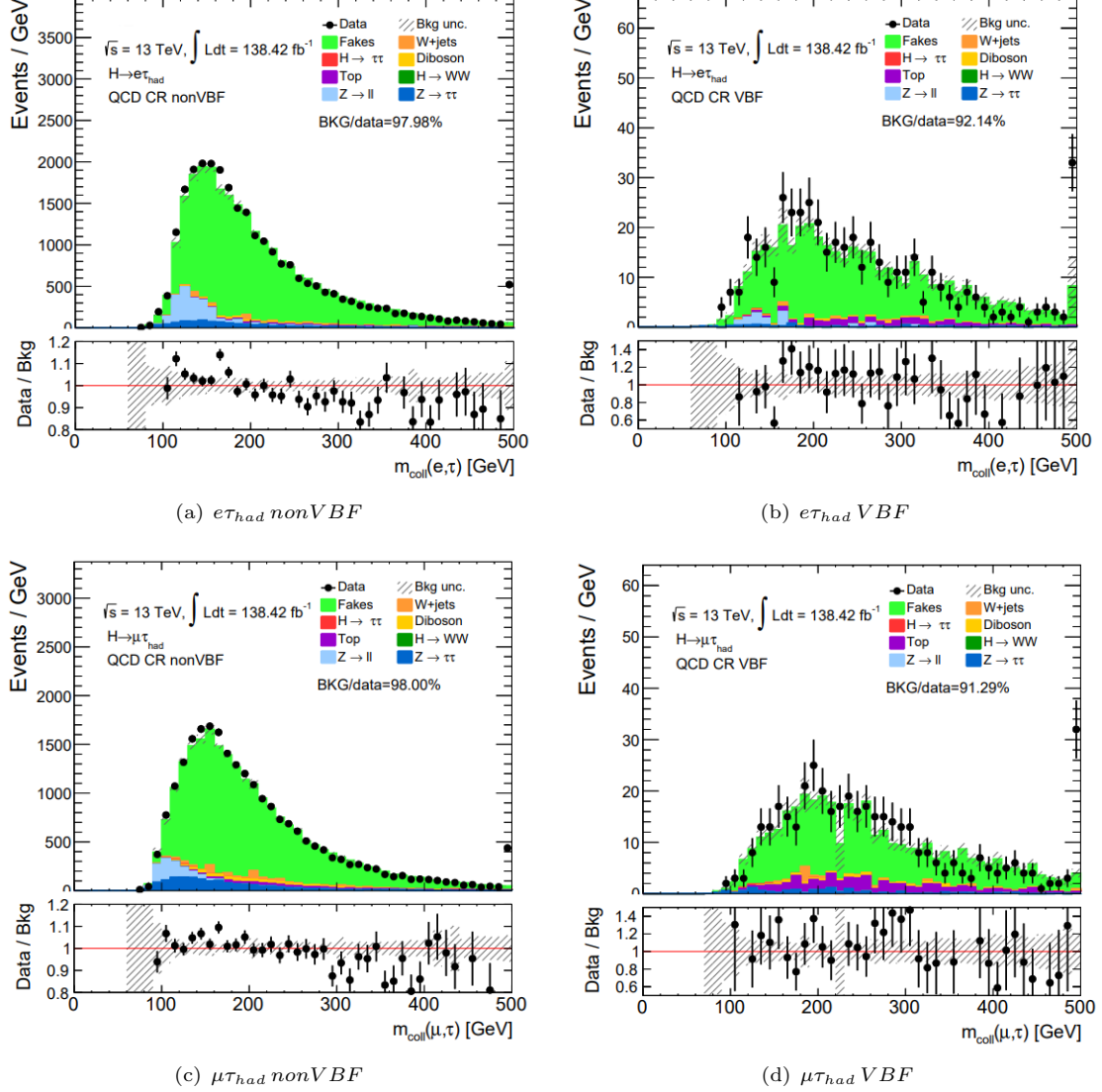


Figure 10.9: The $m_{coll}(\ell, \tau_{had-vis})$ distributions in the QCD control regions, for the (a) nonVBF, $e\tau_{had}$ channel, (b) VBF, $e\tau_{had}$ channel, (c) nonVBF, $\mu\tau_{had}$ channel, (d) VBF, $\mu\tau_{had}$ channel.

10.2 Symmetry Method

In the analysis using symmetry method, only leplep channels are analysed, the studies of the lephad channels are still in progress. The background processes in the leplep channels include the $Z \rightarrow \tau\tau$, diboson, Top-quark production, SM Higgs boson decays ($H \rightarrow \tau\tau$ and $H \rightarrow WW$). The lepton universality in the SM implies that the SM decays involving the electrons and muons are equivalent for these background processes. The expected symmetry is exploited to evaluate these backgrounds, and they are referring to *symmetric background* in the following. In addition, the

misidentified leptons are estimated by two methods as discussed in 10.2.1.1. In lephad channels, the background processes are the same as those introduced in section 10.1.2. In addition to the symmetric backgrounds as mentioned, the W +jets processes are also considered as symmetric under the assumption that the event with a jet misidentified as τ_{had} is symmetric under the e - μ exchange.

10.2.1 $\ell\tau_{\ell'}$ Channel

The background estimation in the leplep channels exploited the data-driven technique. Under the assumption of the $e - \mu$ symmetry, the datasets of the $e\tau_\mu$ channel and the $\mu\tau_e$ channel have the same contribution from the SM background. However, the lepton originate from the Higgs boson is most likely the leading lepton, meaning the $H \rightarrow \ell\tau$ signal is mostly present in the $\ell\tau_{\ell'}$ dataset. Therefore, in a search for the $H \rightarrow e\tau$ decays, as example, the SM background can be estimated using the $\mu\tau_e$ dataset, and vice versa.

This background estimate is corrected with the efficiency corrections for restoring the expected symmetry in the SM events, as discussed in chapter 9. Only the true leptons are efficiency corrected, the contributions of the misidentified leptons need to be subtracted from data before the efficiency corrections, as discussed in section 10.2.1.1. According to equation 9.9, the $e\mu$ datasets are used in the background estimation of the μe channel, and the efficiency corrections should be applied to the $e\mu$ channel as:

Sample		$H \rightarrow e\tau$		$H \rightarrow \mu\tau$	
		nonVBF	VBF	nonVBF	VBF
Signal	ggF	-720 ± 230	-11 ± 4	480 ± 190	8 ± 4
	VBF	-38 ± 12	-20 ± 7	26 ± 11	14 ± 6
	VH	-31 ± 11	-0.55 ± 0.21	21 ± 9	0.33 ± 0.16
	$\widetilde{ggF}$	48 ± 17	1.3 ± 0.5	-35 ± 15	-1.0 ± 0.5
	$\widetilde{VBF}$	3.9 ± 1.3	1.7 ± 0.6	-2.8 ± 1.2	-1.2 ± 0.5
	$\widetilde{VH}$	5.7 ± 2	0.14 ± 0.06	-4.2 ± 1.8	-0.10 ± 0.05
Symmetric	$\widetilde{\text{Data}}$	193700 ± 1200	4300 ± 50	162900 ± 900	3890 ± 50
background	$\widetilde{F_{FF}}$	-19000 ± 2000	-340 ± 40	-15100 ± 1500	-300 ± 34
	$\widetilde{F_{MC}}$	-12600 ± 1800	-112 ± 30	-5400 ± 700	-72 ± 17
Misidentified	F_{FF}	17900 ± 1800	330 ± 40	17200 ± 1700	320 ± 40
background	F_{MC}	5700 ± 800	81 ± 19	11500 ± 1600	106 ± 28
Total background		185700 ± 500	4260 ± 50	171000 ± 500	3950 ± 50
Data		184887	4230	171675	3967

Table 10.4: Summary of the events yields for both channel and category in the symmetry method, after the statistical analysis. The signal event yields are scaled for $\mathcal{B}(H \rightarrow e\tau) = -0.33\%$ and $\mathcal{B}(H \rightarrow \mu\tau) = 0.25\%$. Contributions in the search for $H \rightarrow e\tau$ coming from the $\mu\tau_e$ selection are marked with the "tilde" sign. The statistical and systematic uncertainties are shown [51].

$$(n^{\mu e} - f^{\mu e}) = \mathcal{R}^e(n^{e\mu} - f^{e\mu}) \quad (10.10)$$

where n represents the data events and f represents the fake lepton contributions. Hence, the full background estimate for the μe events $b^{\mu e}$ is:

$$b^{\mu e} = \mathcal{R}^e(n^{e\mu} - f^{e\mu}) + f^{\mu e} \quad (10.11)$$

where $\tilde{n}^{e\mu} = \mathcal{R}^e(n^{e\mu} - f^{e\mu})$ represents the symmetric background estimation. The event yields of the data, signal samples, symmetric background and misidentified background samples are shown in table 10.4. Figure 10.10 shows the representative kinematic distributions for both channel and category, a good modelling can be observed.

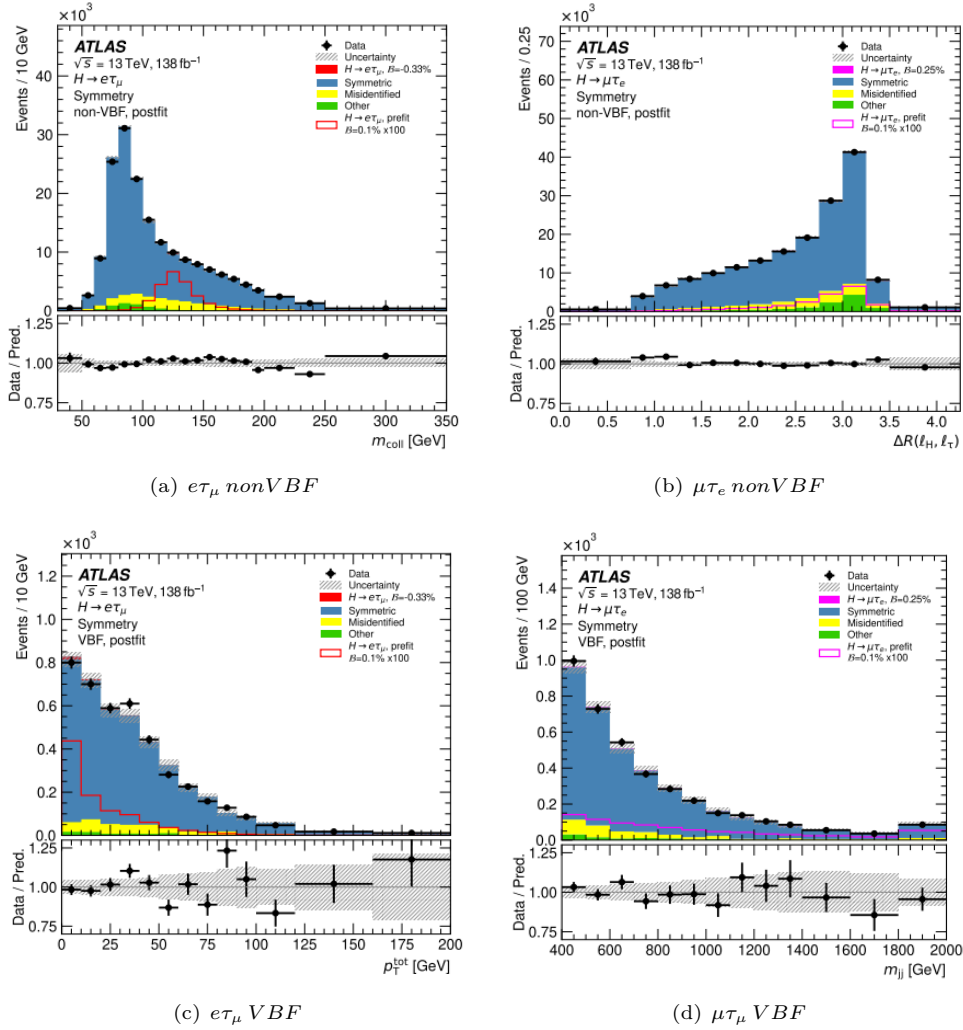


Figure 10.10: Representative kinematic distributions for $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states, and for both nonVBF (top) and VBF (bottom) category. Obtained by performing independent fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals, the chosen kinematic variables are: (a) collinear mass, (b) the angular separation between ℓ_H and ℓ_τ $\Delta R(\ell_H, \ell_\tau)$, (c) the vector sum of p_T of the two leptons and (d) the invariant mass of the dijet system m_{jj} . The hashed band represents the prefit statistical uncertainty, and the experimental and theoretical uncertainties [51].

10.2.1.1 Misidentified Lepton Estimation

The misidentified light leptons are estimated by two methods: the MC simulation and the data-driven fake-factor method, depending on which sources misidentified as the light leptons.

The fake background contributions from the jet misidentified as light leptons and the light leptons originating from hadronic decays within a jet are estimated by the fake-factor method. This approach is similar to the one discussed in section 10.1.2.1 for the lephad channels, but it is used for estimating the misidentified light leptons instead in the lelep channels. Most of the fake contributions originate from the W +jets production, one light lepton originate from the W boson decay, and the jets is misidentified as the other lepton. While a small fraction of contributions originate from the QCD multi-jets processes, both light leptons could be fake by the jets.

Similar to lephad channels, anti-ID and anti-iso regions are defined by inverting the identification and isolation requirements of the light leptons, respectively. A transfer factor or fake factor is defined as the ratio of data events in ID and iso regions over the data events in anti-ID and anti-iso regions:

$$FF = \frac{N^{ID,iso}}{N^{anti-ID,anti-iso}} \quad (10.12)$$

Thus, the number of events with a misidentified lepton in the signal region can be obtained as:

$$N_{fakes}^{SR:ID,iso} = FF \times N^{SR:anti-ID,anti-iso} \quad (10.13)$$

However, detector effects can give rise to different FFs for electrons and muons. The FFs should be binned in the kinematic properties of the leptons and applied as event weights, the kinematic properties are the transverse momentum and the pseudorapidity of the leptons. In addition, both leptons could be faked and double counted in the estimation of the fake background in the signal region, and they should be subtracted:

$$N_{fakes}^{SR} = \sum_{i=1}^{N_e^{SR:anti-(ID,iso)}} FF_e(p_{T_i}^e, \eta_i^e) + \sum_{i=1}^{N_\mu^{SR:anti-(ID,iso)}} FF_\mu(p_{T_i}^\mu, \eta_i^\mu) - \sum_{i=1}^{N_{e,\mu}^{SR:anti-(ID,iso)}} FF_e(p_{T_i}^e, \eta_i^e) \cdot FF_\mu(p_{T_i}^\mu, \eta_i^\mu) \quad (10.14)$$

The fake factors are measured in a dedicated data-driven Z +jets CR which enriched in misidentified light leptons. Events in the Z +jets CR should have exactly three leptons, two of which with opposite-sign and same flavour. and they are required to be consistent with the Z boson mass peak. The third lepton is known as the *probe* lepton, which is mostly originated from a jet misidentified as lepton.

The fake background contributions from τ_{had} misidentified as light lepton, the muons misidentified as electrons and photon misidentified as electrons are estimated from the MC simulations. The simulated events are selected according to the baseline selection, and the simulation of the following processes are considered: $Z \rightarrow \tau\tau$, Top-quark production, diboson, $Z \rightarrow \ell\ell$ and vector boson associated with γ . The largest contribution to the fake background among these processes come from the $Z \rightarrow \mu\mu$ processes, where muons with low p_T could be misidentified as electrons. Dedicated $Z \rightarrow \mu\mu$ validation region is defined to check the modelling of this process.

10.2.2 $\ell\tau_{had}$ Channel

The ongoing studies of the $\ell\tau_{had}$ channels are also based on the assumption that the symmetry in the SM processes under the exchange of prompt electrons with prompt muons is applied to a

good approximation, any signal of the lepton-flavour-violating decays of the Higgs boson break this symmetry.

The analysis also employed both MC simulation and data-driven techniques to estimate the backgrounds. The effects of the detector are taken into account through efficiency corrections. The correction strategy for the lephad channels is discussed in Chapter 9. The full background estimation follows equation 10.11.

In the $\ell\tau_{had}$ channels, the background processes considered as symmetric background include the $Z \rightarrow \tau\tau$, production of dibosons, the production of top quarks, the W +jets in which the jet is faking the τ -lepton¹, and the decays of the Standard Model Higgs boson into tau leptons ($H \rightarrow \tau\tau$) and into pairs of W bosons ($H \rightarrow WW$).

The contribution of jets-faking-tau is estimated using the fake-factor method, as explained in section 10.1.2.1. On the other hand, the misidentified lepton background is estimated using the ABCD method, as discussed in section 10.1.1.1. At the time of writing, an unexpected asymmetry was found in the W +jets fake distribution, which refers to the process of a W boson and a jet-faking-tau. Therefore, another approach was considered to count the contribution of misidentified lepton in equation 10.11², which means the background estimation follows:

$$b^{\mu\tau} = \mathcal{R}^e(n^{e\tau} - f_{fake\ lepton}^{e\tau} - f_{MC\ fakes}^{e\tau}) + f_{fake\ lepton}^{\mu\tau} + f_{MC\ fakes}^{\mu\tau} \quad (10.15)$$

where MC fakes refer to the fake background contributions originate from lepton-faking-tau, lepton-faking-lepton, and the processes of W +jets (including a W boson plus a jet-faking-lepton). While fake lepton contributions are referring to the jet-faking-lepton and estimated by the data-driven technique, which includes the contributions originate from QCD multi-jets processes and Z +jets processes, and in the following, "other fakes" denote the fake backgrounds other than these jet-faking-lepton background contributions. The jet-faking-lepton estimation exploits the ABCD method with a lepton fake-factor (LFF), which is defined as:

$$LFF = \frac{N_{Data,SS}^{ID,Iso} - N_{MC,true\ lepton+other\ fakes,SS}^{ID,Iso}}{N_{Data,SS}^{ID,anti-Iso} - N_{MC,true\ lepton+other\ fakes,SS}^{ID,anti-Iso}} \quad (10.16)$$

where SS represents same sign. The jet-faking-lepton contributions are estimated by the following equation which takes the ratio of isolation (Iso) to anti-isolation (anti-Iso):

$$N_{FakeLep} = (N_{Data,OS}^{ID,anti-Iso} - N_{MC,true\ lepton+other\ fakes,OS}^{ID,anti-Iso}) \times LFF \quad (10.17)$$

where OS represents opposite sign and LFF is estimated by equation 10.16.

Figure 10.11 shows distributions of the collinear mass and transverse mass before the statistical fit for the $e\tau_{had}$ and $\mu\tau_{had}$ final states. A good agreement between the data and background predictions is observed in these kinematic distributions. Only the statistical uncertainty is included and the systematic uncertainties have not been considered in these distributions.

¹For those W +jets events in which the jet is faking the electron or muon, they are included in the MC fakes estimation.

²The jet-faking-tau background originate from the W +jets process is expected to be symmetric and therefore, is included in the symmetric background instead, while the contribution from QCD multi-jets is counted by the fake lepton background.

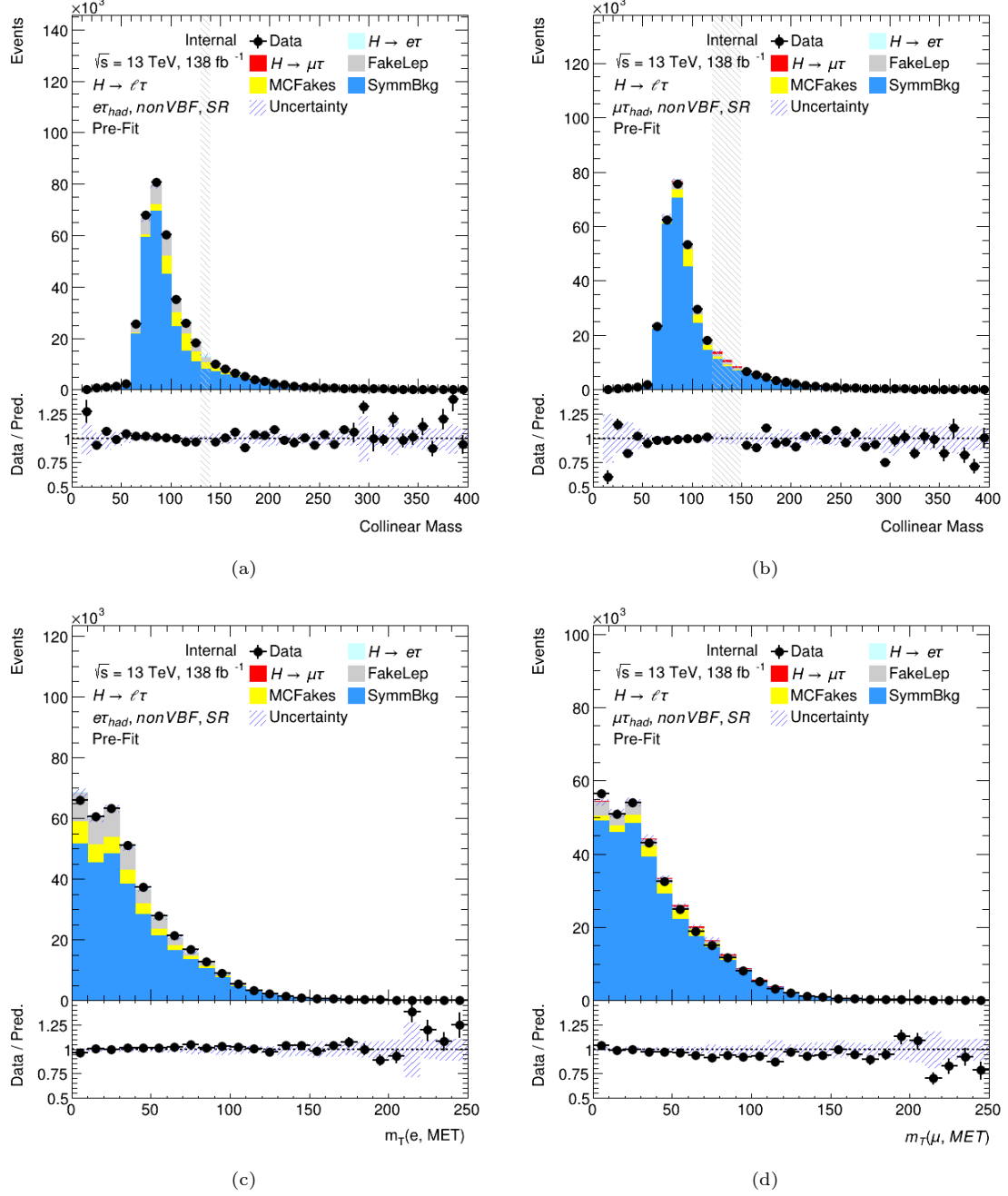


Figure 10.11: The pre-fit distributions of the collinear mass (Top) and $m_T(\ell, E_T^{miss})$ (bottom) for the $e\tau_{had}$ (left) and $\mu\tau_{had}$ (right) final states. Only the statistical uncertainty is included.

Chapter 11

Multivariate Analysis and Machine Learning

The datasets collected from the modern high-energy-physics experiments have very high complexity and dimension, the LHC produced gigantic data outputs from the proton-proton collisions during its operational periods. Traditional data analysis such as CBA is no longer able to be used for analysing the high-dimensional complex datasets with sufficiently high accuracy. In addition, the developments on the computing power also motivated particle physicists to analyse the data by employing algorithms that utilize multiple variables simultaneously. The multivariate analysis (MVA) exploits the machine learning techniques, such as the artificial neural networks and decision trees.

Machine learning techniques are exploited in wide variety of applications and challenges. There are two basic approaches: the supervised learning and unsupervised learning, the supervised learning uses the labeled data and the latter does not. The supervised learning approach can be used in two main types of tasks: the classification and regression. The classification refers to the problem of identifying an observation belong to which categories, such as separating signal from backgrounds. While the regression refers to the problem of understanding the relationship between the dependent and independent variables, which are very useful in the prediction and forecasting tasks. In some tasks, the unlabeled data need to be analysed and clustered without any supervision, which can be done by using the unsupervised approach. There are three main types of tasks in unsupervised learning: clustering, association and dimensionality reduction. The clustering refers to the problem of grouping unlabeled data according to their features. The association is used to relate each variables in a given dataset and measure their dependencies. While the dimensional reduction is used to reduce the dimension of a dataset with too many features.

The dataset generated from the LHC have very high dimensionality and large volume, the machine learning techniques are widely used in many areas for handling the overwhelming information. At the detector level, machine learning techniques can be employed to monitor the data quality and detector performance automatically. Machine learning is also widely used in the event identification, reconstruction and classification tasks, such as increasing the accuracy of the particle tagging (for example, the b-tagging) [144], pile-up mitigation [145], objection identification and reconstruction (for example, the RNN tau ID) [124], supervised event classification and many others. In the physics analyses, machine learning is used to search for the rare signals among the vast backgrounds and

achieving a high signal sensitivity.

In the Higgs LFV analysis, machine learning techniques based on the neural network and boosted decision trees are employed for enhancing the signal-to-background separation. The BDT technique is used in the MC-template method, while fully connected deep neural networks are exploited in symmetry method.

In this chapter, the machine learning techniques that are used in the studies of the lephad channels and MC-template method are introduced. Although the neural network algorithm is not used in the MC-template method, its performance in each channel and category have been tested. A good performance is achieved by using a full connected deep neural network in the symmetry method, which could be potentially shown a better performance in the Run-3 analyses, the neural network algorithm is introduced in section 11.1. The BDT algorithm, exploited in the MC-template method, is discussed in section 11.2. Different channels and categories exploited different training strategies, the details are discussed in section 11.3.

11.1 Neural Network Algorithm

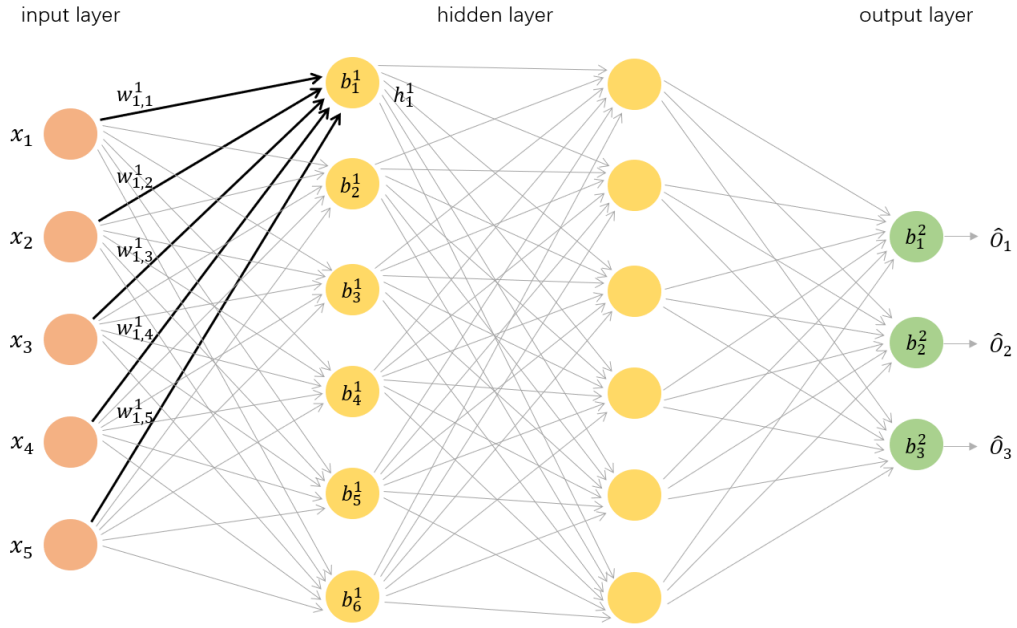


Figure 11.1: The architecture of an example neural network with four layers.

The basic building block of a neural network is the neurons, which take the inputs to the network and give the outputs after some calculations. One neural network can be built by connecting the neurons in each layer, figure 11.1 shows the architecture of an example neural network, the five neurons or nodes in the first layer (the input layer) corresponding to the five input variables, the nodes in the hidden layers and the output layers with three nodes corresponding to three possible classes.

The algorithm is to multiply each input by a weight w , and sum up all the weighted inputs with a bias b , the sum is then input to an activation function for getting a new input to the next layer. For example, $h_1^1 = f(\mathbf{w}_1^1 \mathbf{x} + b_1^1)$, where $\mathbf{x}$ is a vector with all the input values, the h_1^1 is an

input to the next layer and repeat the calculation until getting the output $\hat{O}_i$, where $i = 1, 2, 3$, as shown in figure 11.1. In general, the input variables and the layer structure (for example, the number of neurons in each layer) should be optimized in such neural network in order to obtain the desired accuracy.

The weight w and the biases b are determined by training the neural network. The training process is to use a function, called the *loss function*, mapping the an event of multiple variables onto a real number and evaluate how well the algorithm models the dataset. The training is to minimize the loss function. In a classification task, the most general loss function is the *categorical cross-entropy*:

$$L = -\sum_{i=1}^N y_i \cdot \log \hat{y}_i \quad (11.1)$$

where N is the number of events, y_i represents the i -th element of the true label vector y and $\hat{y}_i$ is the i -th element of the predicted probability. The loss function implies that the training is better with larger number of events.

The symmetry method (lelep channels) exploits a fully connected deep neural network, which achieved a good classification performance. There are many other different types of neural networks, such as the recurrent neural network used in the tau identification, the graph neural network used in particle tracking and reconstruction and many others, allowing for more optional functions and features to be used in the training. However, it was found that the NN training performance of the dataset of the MC-template method is nearly the same as the BDT training performance. In comparison, BDT has a simpler structure and is easier to be built up. Therefore, the MC-template method exploited the BDT technique. The NN trainings could potentially have a better performance than the current BDT methods in the Run-3 and HL-LHC experiments as more statistics are allowed.

11.2 Boosted Decision Tree Algorithm

Decision tree learning is a tree-like model of decisions that used in supervised learning. Classification decision tree is a classifier that has a binary structure, figure 11.2 shows the architecture of a single binary decision tree. The decision trees are used to categorize the events from a sample as either signal or background. An event enter this flowchart begin with the root node, while each internal node tests the event based on a single discriminating variable to decide how likely the event is a signal or background. Decision tree is formed with the leave nodes which represent the class labels, and the path of the event from the root to the leaf represents a full classification process in the training.

The training of the events start from the root node, and a sequence of binary splits based on the discriminating variables and their cut criteria that have the largest separation between signal and background is applied to the events. The cut criteria of each node are determined by training the decision tree, the splits stop once a certain node has satisfied the certain criteria such as minimum or maximum number of events or signal purity (purity = $S/(S + B)$). These nodes are then defined as the leave nodes labeled as signal or background.

The meaning of boosting is that each tree is dependent on the prior trees. A boosted decision tree is an ensemble learning method that each tree corrects the errors of previous trees, and the predictions are drawn based on the ensemble of trees. There are two common boosting algorithms:

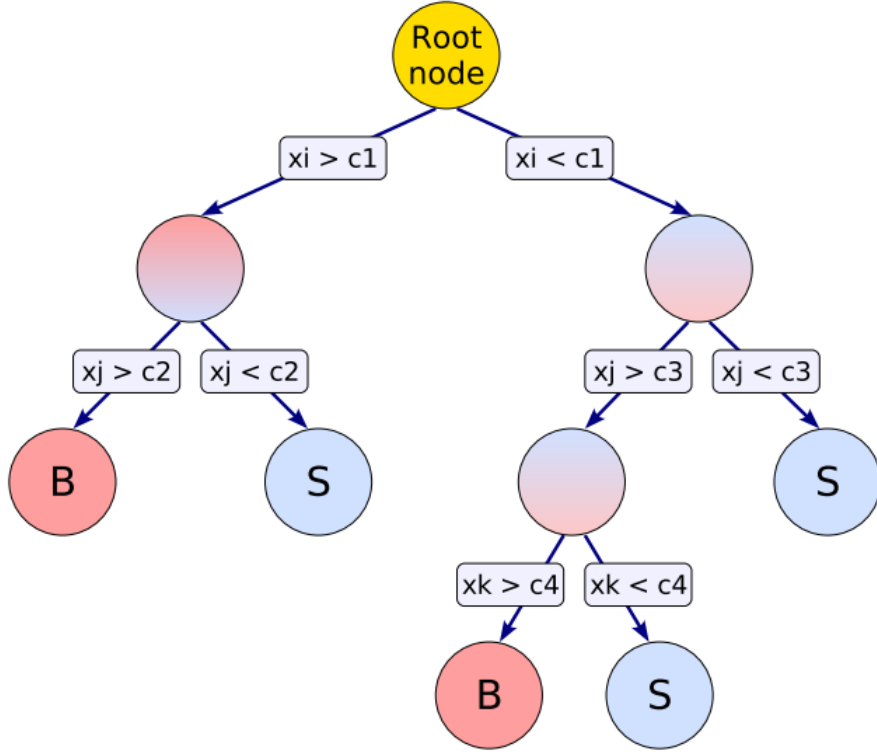


Figure 11.2: Schematic view of a binary decision tree. The x_i represents the event values, c_i represents the corresponding cut criteria. In the leaf node, the "S" represents a signal class and the "B" represents the background class [146].

the adaptive boosting and gradient boosting. In the adaptive boosting algorithm, signal (background) events misidentified as background (signal) are given a higher weight than events that are classified as background (signal) in the leaf node. A new decision tree is trained with the re-weighted events, the boosting is iterated and a *forest* is formed with all the decision trees. While for the latter algorithm known as the gradient boosting, each tree is trained for minimizing the loss function of the previous tree using gradient descent. The boosting is iterated, and the gradient of the loss function is computed with respect to the predictions of the current ensemble of trees, a new decision tree is trained for minimizing this gradient, and then the new tree is included in the ensemble of trees. The iteration is stopped until the stopping criterion is satisfied. In addition to the gradient boosting, an optional procedure can be implied simultaneously to increase the accuracy of the trainings, which is the *bagging* procedure known as the *stochastic gradient boosting*. The bagging procedure uses a random fraction of the training events in the training set, while the rest of the events are used for the validation set. The main drawback of this procedure is the increased training time.

11.2.1 Hyperparameters

In order to optimize the performance of the BDT training, the hyperparameters that define the architecture of the BDT should be optimized. The hyperparameters are a set of parameters which constrain the learning process, they have to be tuned for getting the optimal model which minimizes the loss function.

During the tuning, the risks of *overfitting* must be considered. Overfitting issues occur when

machine learning model provides accurate predictions for the training data but not for new data, accurate predictions are achieved only if all types of data is generalized within its domain. The overfitting issue in a machine learning model can be detected by using the *k-fold cross-validation*, which is discussed in section 11.2.2.

There are several main hyperparameters of the BDT training need to be optimized, depending on the features of the input events. They are introduced and listed as follow:

- Number of trees: The total number of trees to create in the ensemble. In general, a larger forest can convert more information and achieve a higher accuracy of the predictions. However, the model is very easy to be overfitted and the performance will be saturated.
- Maximum depth: The longest path from the root node to the leave node. Deeper trees have more splits and captures more information about the events. If the depth is too high, the model will be very easy to be overfitted and the performance will be saturated.
- Minimum node size: The minimum number of observations in a node. This parameter is a kind of stopping criterion as the node will not be further split if one node with smaller observations, and the node becomes the leaf node.
- Shrinkage: The learning rate of gradient boosted decision tree.

11.2.2 Cross-validation

Not all the data are used for the training set in a machine learning model, a test set should be defined and used for evaluating the training performance of the classifier. However, there is a risk of overfitting on the test set due to the problematic hyperparameters, and the knowledge from the overfitted test set can affect the model. In order to solve this problem, an independent dataset is taken from the data and used for validation. After the training of the training set and before the final performance evaluation on the test set, additional evaluation should be done on the validation set.

Splitting the data into three sets will reduce the statistics that can be used for learning the model, and it is unreliable to pick the datasets randomly. A approach which is known as k-fold cross-validation is exploited to solve this problem. In this approach, the data is split into k parts. The first model is trained with the $k - 2$ folds as the training data, then the $k - 1$ fold is used for validation, and the remaining fold is used for test set. The process is repeated k times until each of the k folds has been used as test set. The common choices of the k value is 5 or 10 in machine learning models.

11.2.3 Performance

After learning the model, in order to represent how likely an event is a signal or background, a score is assigned to each event. The events are distributed within a score range from -1 to 1 (from 0 to 1 for NN score) according to their BDT scores, and events with BDT score close to 1 (-1) is more likely a signal (background). Base on the BDT score distribution, the performance of the classifiers can be quantified as the binned significance Z , which is defined as:

$$Z = \sqrt{\sum_i Z_i^2}, \quad Z_i = \sqrt{2[(s_i + b_i)\log(1 + \frac{s_i}{b_i}) - s_i]} \quad (11.2)$$

where s_i and b_i represent the number of signal and background events in the i^{th} bin, respectively. The BDT score distributions with higher binned significance have larger separation between the signal and background.

In addition, the discrimination performance of a BDT training can be represented by the *receiver operating characteristic curve* (ROC curve). Which shows the relation between the signal efficiency (or true positive rate) and background rejection (or false positive rate) for the classifiers. The area under the curve (AUC) is a quantity to represent the performance of the classifier, and a classifier with higher signal efficiency and background rejection (larger AUC) have a larger separation power between the signal and background.

11.3 Training Strategies

In the $\ell\tau_{had}$ analysis, the MVA technique based on BDT is implemented via the Toolkit for Multivariate Data Analysis (TMVA) [146], the gradient boosting and 5-fold cross-validation were used. Different BDT trainings are performed for the $\mu\tau_{had}$ and $e\tau_{had}$ channels, as well as for the VBF and nonVBF regions. Instead of rejecting all the background processes by using one BDT, different BDTs are trained focusing on different backgrounds, depending on the statistics. The specific features of different backgrounds can be exploited to improve background efficiency by using this strategy. In this approach, two or three BDTs are trained focusing on different set of backgrounds. Each BDT provides a BDT score for each event, the scores are combined into one via linear combination or quadratic combination:

$$S_{linear}^{comb} = \frac{1}{\sum_i c_i} \sum_i c_i S_i, \quad S_{quadratic}^{comb} = \frac{1}{\sqrt{\sum_i c_i^2}} \sqrt{\sum_i c_i^2 S_i^2} \quad (11.3)$$

where S_i represent the score obtained with the i classifier and c_i is the combination coefficient of the i classifier. The combined BDT score distributions are used as final discriminant.

There are two options to split the backgrounds for trainings in the $\ell\tau_{had}$ analysis, base on the two main backgrounds: the $Z \rightarrow \tau\tau$ and misidentified $\tau_{had-vis}$. The first option is to discriminate the signal from the $Z \rightarrow \tau\tau$ events by training the first BDT (BDT₁), while the second BDT (BDT₂) is trained to separate the signal from the rest of the backgrounds. In the second option, BDT₁ is trained to separate the signal from the $Z \rightarrow \tau\tau$ events and BDT₂ separates the signal from the misidentified background, while the BDT₃ is trained to separate the rest of the backgrounds. Different options and combination methods are chosen independently for each channel and category, depending on which one can provide highest binned significance. The combination coefficients are optimized by performing grid searches. The combinations which provide the highest binned significance for each channel and category are summarised in table 11.1. Each of the BDTs in all the channels and categories are shown in figure 11.3 and figure 11.4.

The hyperparameters as listed in section 11.2.1 were optimized separately for each channel and category, and for each BDT as well. Grid search approach was used for the hyperparameter optimization and based on the BDT which was trained against all backgrounds. A quick re-optimization was performed on each BDT, no hyperparameter with better performance was found. However, in the $\mu\tau_{had}$ VBF region, a set of hyperparameters with significantly better performance was found. The summary of the optimized hyperparameters for each channel and category is shown in table 11.2, with additional information on other BDT configuration options. Gradient boosting is

exploited with the bagging approach for increasing the accuracy. For those options are not shown in table 11.2, they follow the default values in the TMVA user-guide [146].

Region	Channel	Option	Coefficients	Combination
nonVBF	$e\tau_{had}$	3 BDTs	$c_1 = 0.1, c_2 = 0.4, c_3 = 0.1$	Linear
	$\mu\tau_{had}$	2 BDTs	$c_1 = 0.3, c_2 = 0.8$	Linear
VBF	$e\tau_{had}$	2 BDTs	$c_1 = 0.1, c_2 = 0.4$	Quadratic
	$\mu\tau_{had}$	2 BDTs	$c_1 = 0.1, c_2 = 0.9$	Quadratic

Table 11.1: The summary of the best combinations of the BDT trainings for all channels and regions.

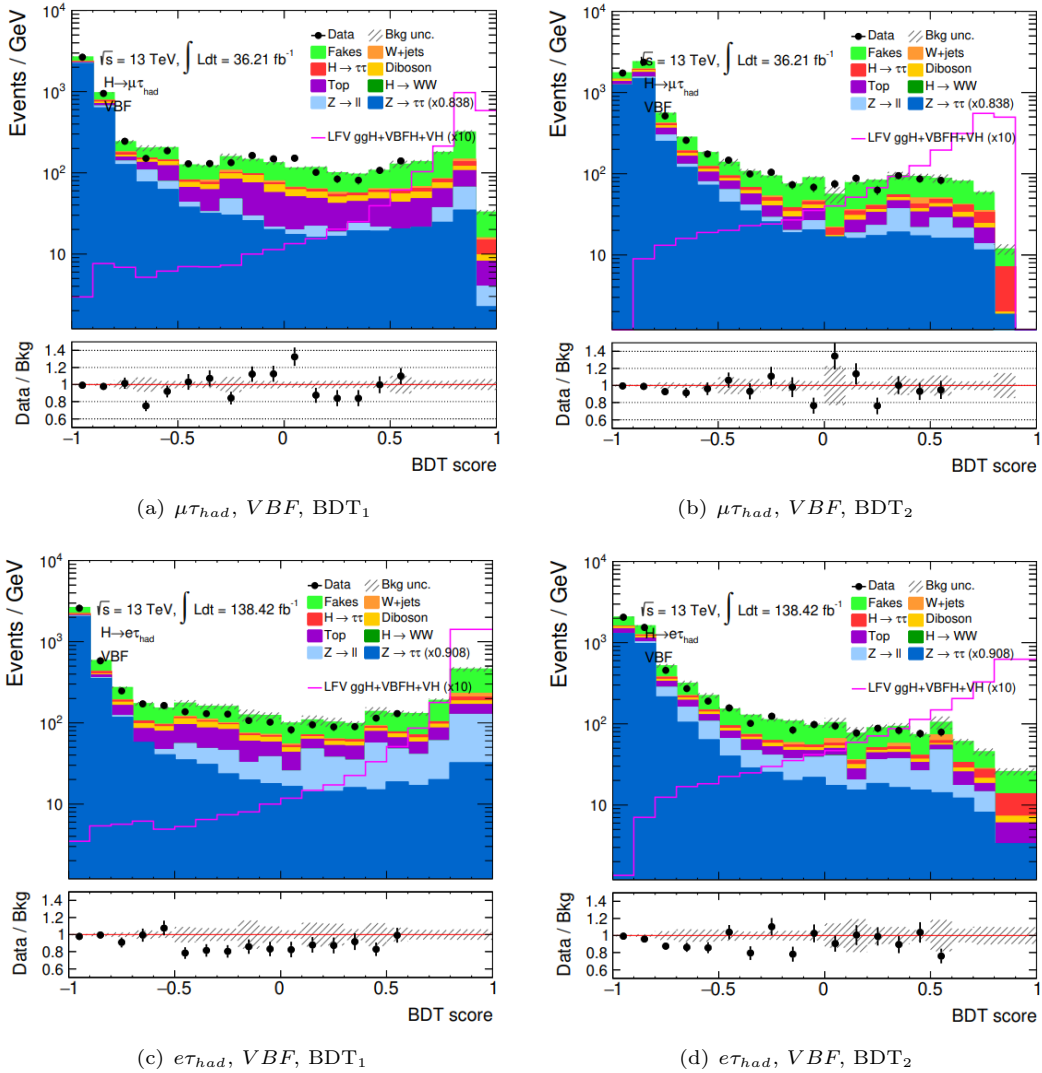


Figure 11.3: The BDT score distributions for the VBF region: (a) $\mu\tau_{had}$ BDT₁ with $c_1 = 0.1$, (b) $\mu\tau_{had}$ BDT₂ with $c_2 = 0.9$, (c) $e\tau_{had}$ BDT₁ with $c_1 = 0.1$ and (d) $e\tau_{had}$ BDT₂ with $c_2 = 0.4$. The $Z \rightarrow \tau\tau$ background in each distribution is scaled by the $Z \rightarrow \tau\tau$ normalization factors calculated from the combined score distributions. Only the statistical uncertainties are shown.

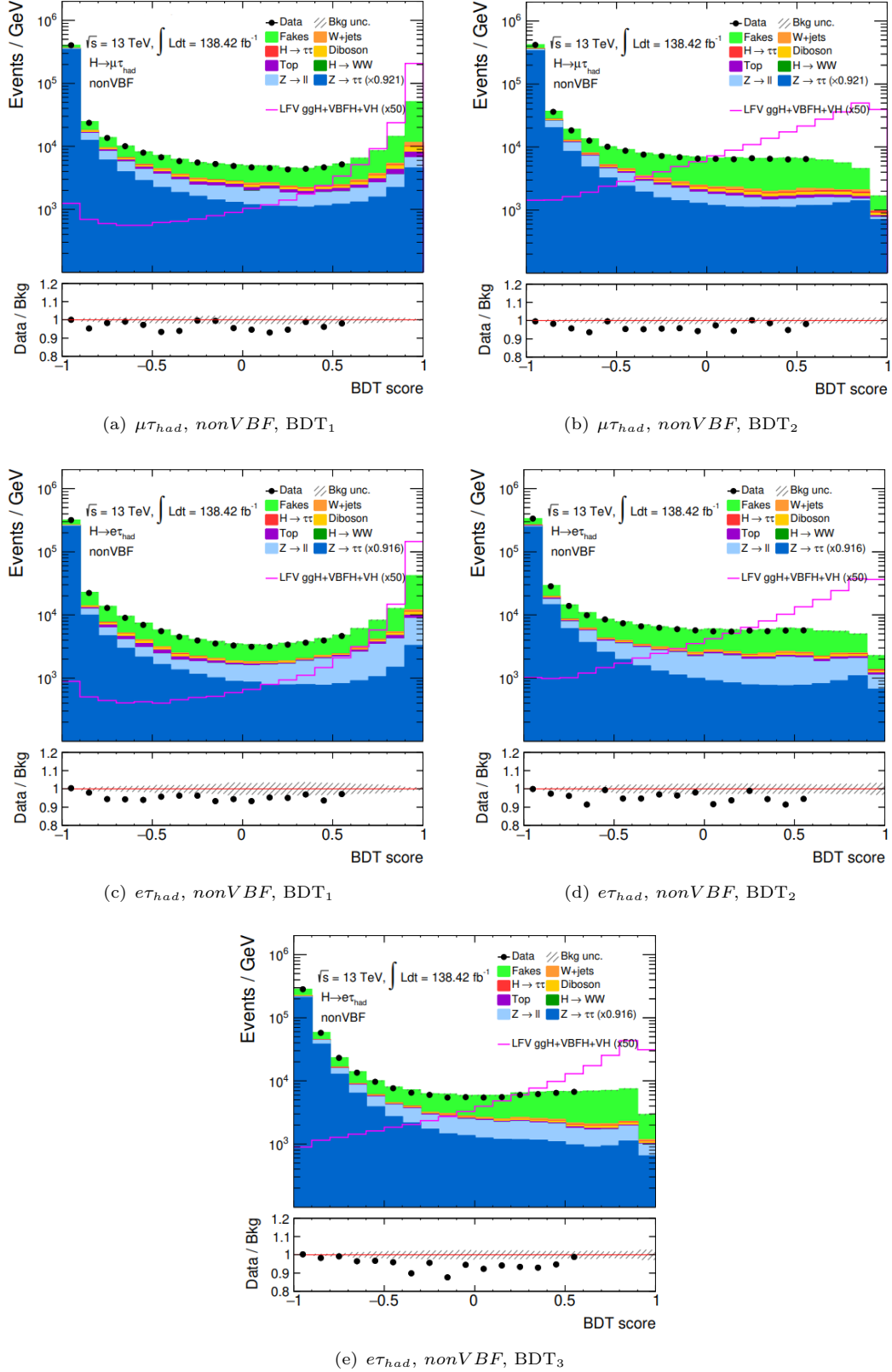


Figure 11.4: The BDT score distributions for the nonVBF region: (a) $\mu\tau_{had}$ BDT₁ with $c_1 = 0.3$, (b) $\mu\tau_{had}$ BDT₂ with $c_2 = 0.8$, (c) $e\tau_{had}$ BDT₁ with $c_1 = 0.1$, (d) $e\tau_{had}$ BDT₂ with $c_2 = 0.4$ and (e) $e\tau_{had}$ BDT₃ with $c_3 = 0.1$. The $Z \rightarrow \tau\tau$ background in each distribution is scaled by the $Z \rightarrow \tau\tau$ normalization factors calculated from the combined score distributions. Only the statistical uncertainties are shown.

Region	Channel	N_{trees}	MaxDepth	MinNodeSize	Shrinkage
nonVBF	$e\tau_{had}$	500	7	1	0.05
	$\mu\tau_{had}$	500	7	1	0.05
VBF	$e\tau_{had}$	300	10	1	0.01
	$\mu\tau_{had}$ BDT ₁	300	8	1	0.009
	$\mu\tau_{had}$ BDT ₂	300	6	1	0.0095
BDT configuration options					
BoostType=Grad		UseBaggedBoost=true		BaggedSampleFraction=0.5	
				nCuts=20	

Table 11.2: Summary of the optimized BDT hyperparameters for all regions and channels. The BDT configuration options for the type of boosting, the boolean option of the gradient boosting with bagging, the fraction of samples used for bagging and the number of steps during node cut optimisation. The other configuration options follow the default values in the TMVA user-guide [146].

The input variables were optimized using the BDT trained against all backgrounds, the details are discussed in section 11.3.1. After all the optimizations, the BDTs are combined into one with the coefficients as shown in table 11.1. The combined BDT score distributions are shown in figure 11.5, and the binned significances obtained from them are summarised in table 11.3, the binned significances for each of the BDTs are also shown.

Figure 11.5 represent the distributions of the BDT scores within the range from -1 to 1 with 20 bins in total. However, the binning of the score distributions have not been optimized for having the best shape. The strategy to rebin the combined score distributions is to start 100 equivalent bins, sub-bins are merged into a new bin starting from the bin with highest score, and met the following criteria:

- The yields of the major backgrounds should be positive in each bin for merging bins dominant by events with negative weight.
- At least 5 background events in each bin, as the last bin may be empty.
- Total background events should be increasing from the right to the left of the distribution.
- The statistical uncertainties for the major backgrounds should be less than 100% for smoothing the fluctuation in the data over background ratio.

The rebinned combined BDT score distributions for all channels and regions are shown in figure 11.6. The binned significances for the $e\tau_{had}$ channel in nonVBF and VBF regions are obtained as 21.63 and 10.63, respectively. While the $\mu\tau_{had}$ channel in the nonVBF and VBF regions are obtained as 28.38 and 11.36, respectively.

11.3.1 Input Variables Optimisation

Optimization of the input variables are carried out using the BDT trained against all backgrounds. The optimization started with about 20 variables, including high-level variables: mass variables as introduced in section 7.3.1, the missing transverse energy E_T^{miss} , the angular separation ΔR , the dijet-system variables and some others. While low-level variables such as the transverse momentum

	VBF		nonVBF	
	$\mu\tau_{had}$	$e\tau_{had}$	$\mu\tau_{had}$	$e\tau_{had}$
BDT ₁ $Z \rightarrow \tau\tau$	7.9	6.52	18.46	14.40
BDT ₂ Fakes+Others	12.14	11.17	27.03	-
BDT ₂ Fakes	-	-	-	20.35
BDT ₃ Others	-	-	-	17.35
Combination	12.79	11.18	27.16	20.72

Table 11.3: Binned significances obtained by each of the BDTs and by combining the scores, in both region and channel. Systematic uncertainties are not included in the significance calculation.

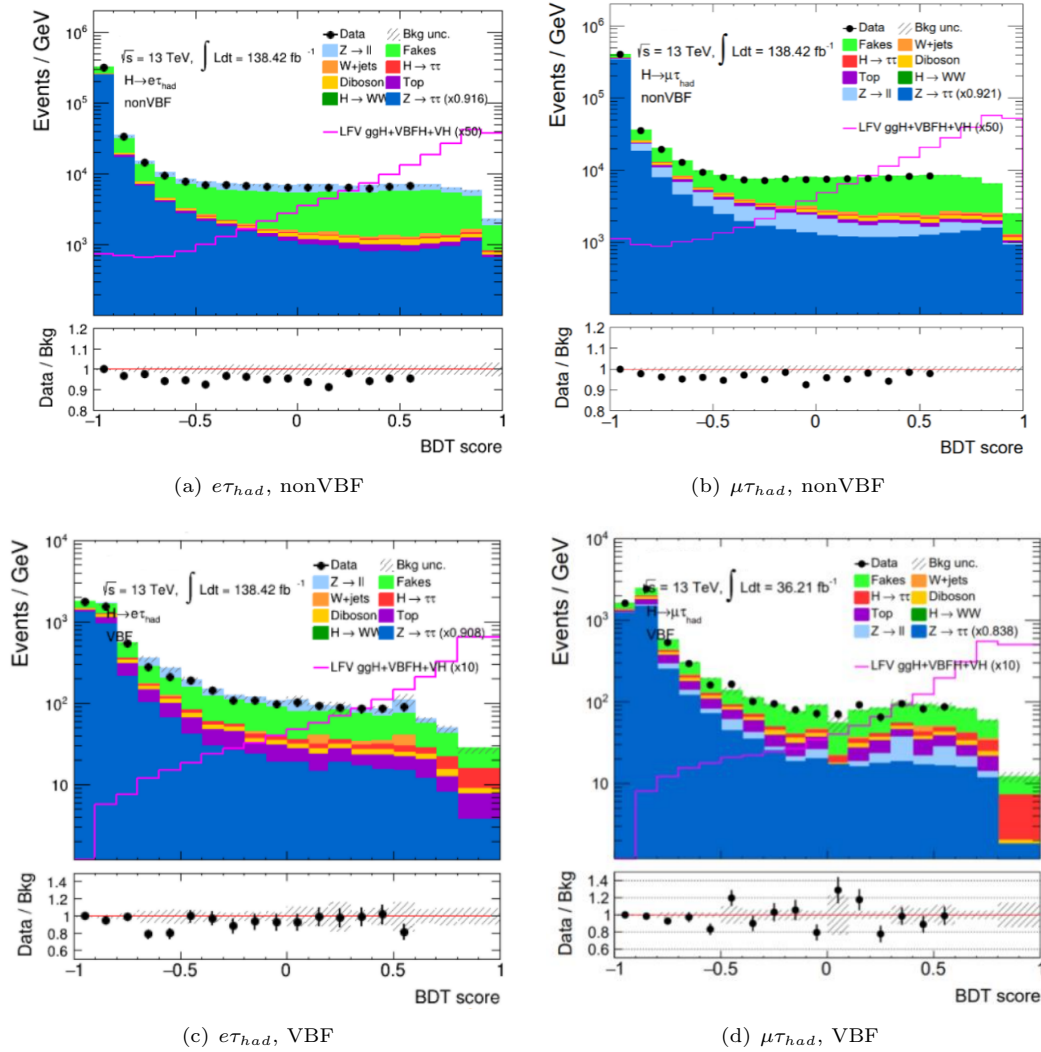


Figure 11.5: The combined BDT score distributions in the nonVBF (top) and VBF (bottom) signal regions, for $e\tau_{had}$ (left) and $\mu\tau_{had}$ (right) channels. The $Z \rightarrow \tau\tau$ backgrounds are scaled by the corresponding $Z \rightarrow \tau\tau$ normalization factors as shown in the legends. Only statistical uncertainties are shown.

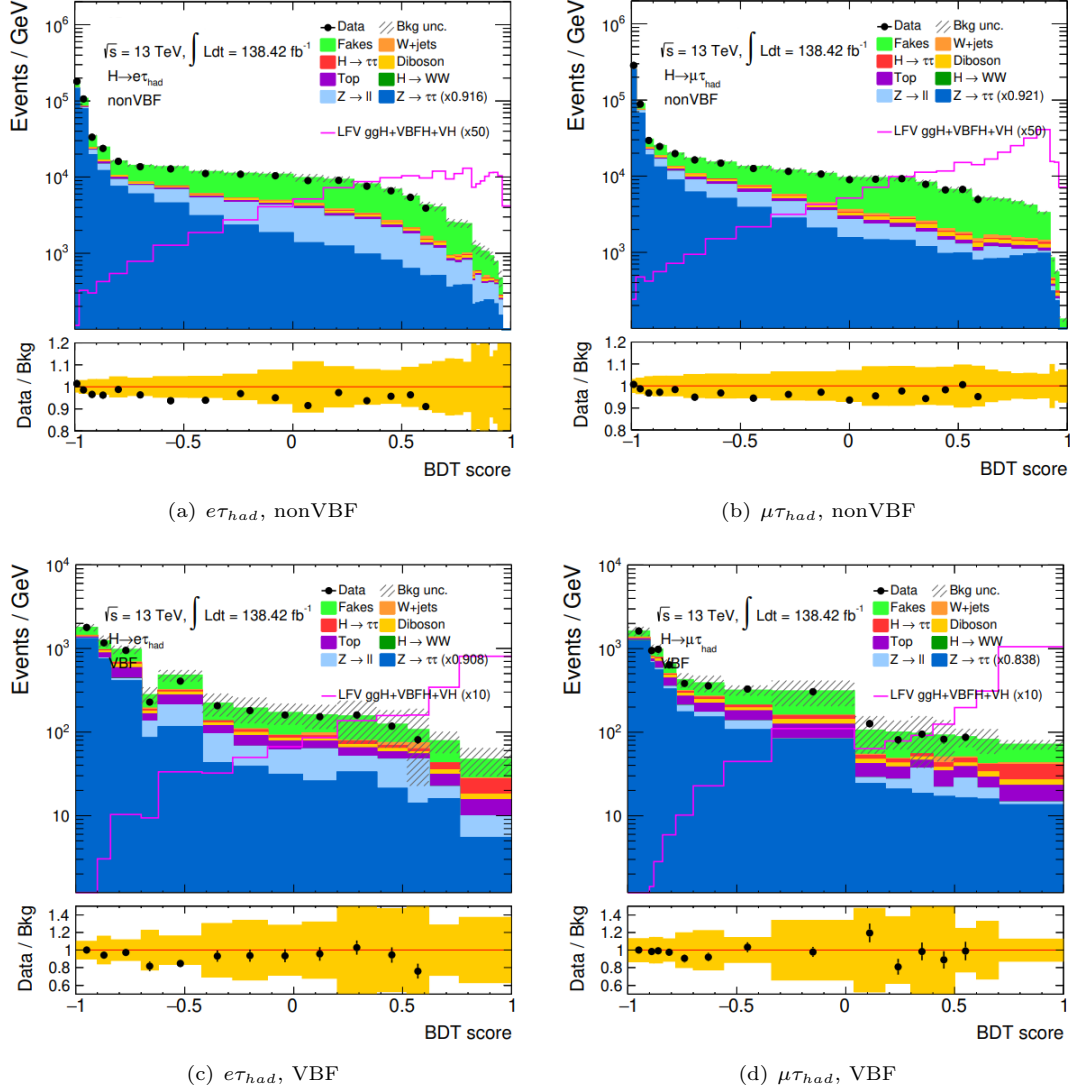


Figure 11.6: The rebinned combined BDT score distributions in the nonVBF (top) and VBF (bottom) signal regions, for $e\tau_{had}$ (left) and $\mu\tau_{had}$ (right) channels. The $Z \rightarrow \tau\tau$ backgrounds are scaled by the corresponding $Z \rightarrow \tau\tau$ normalization factors as shown in the legends. The statistical uncertainties plus a subset of systematic uncertainties are shown.

p_T , the pseudorapidity η and azimuthal angle ϕ , are also considered and utilized to learn the high-level variables and the model. The performance of different combination of these variables is based on the quantity of binned significance. The correlation between each variable and the importance ranking are used as reference for the optimization. The correlation matrix can provide the information of the linear relationship between each variables, when two or more variables are highly correlated, they carry very similar information and should be redundant. Reducing the highly correlated variables can increase the training efficiency and the accuracy of the predictions. While the importance ranking can provide the information of the ranking of the separation power of the input variables. Both correlation matrix and importance ranking are generated by the algorithms implemented in the TMVA [146]. The correlation matrices between the input variables for signal and background, and for each of the BDT in the VBF region $\mu\tau_{had}$ channel, are shown in figure 11.7. In overall, the correlation between each variable is below an acceptable level, as well as the

other channels and regions.

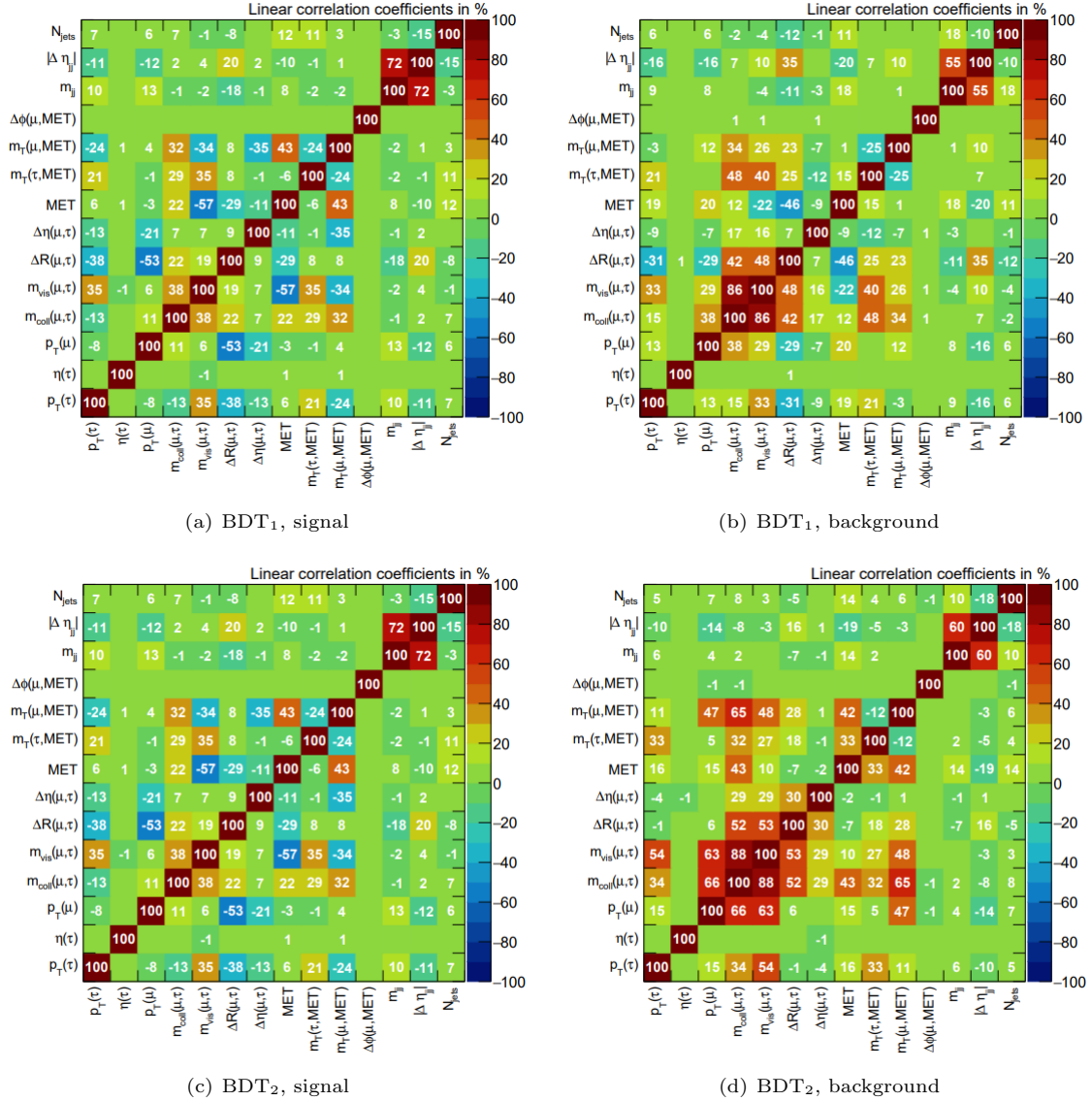


Figure 11.7: The correlation matrices between the input variables for signal (left) and background (right) events, and for BDT1 (top) and BDT2 (bottom) in the VBF region, $\mu\tau_{had}$ channel. The correlation matrices were generated by using the algorithms in TMVA [146].

During the optimization process of the BDT input variables for the $\mu\tau_{had}$ VBF region, a significant improvement in the binned significance after including the variable $N_{jets(p_T > 30)}$ is observed. The binned significances obtained by the trainings with and without this variable are 13.62 and 15.98¹, respectively. Therefore, this variable is motivated to be included. However, if the main contribution comes from the distinction between 2 versus more jets, that could be problematic due to the mis-modelling of the events with more than 3 jets. If the discrimination power of this variable was coming from the distinction between 2 versus 3 jets, then it is safe to include this variable.

The investigation of the variable $N_{jets(p_T > 30)}$ is started by excluding the possibility that the power was coming from the randomness of the BDT training, by checking performance of several

¹The binned significances were obtained by using different samples compare to those used in the final results, but with similar performance.

random seeds used in the BDT trainings with and without $N_{jets(p_T>30)}$. The obtained binned significances are very closed to the original results, the effects from the randomness was excluded.

In order to test whether the discrimination power of $N_{jets(p_T>30)}$ was coming from the distinction between 2 versus 3 jets or 2 versus more jets, a new boolean variable is defined to check the distinction between 2 versus 3 jets:

- If $N_{jets(p_T>30)} = 2$, boolean variable=true (false otherwise).

This boolean variable can be used to test the distinction from 2 versus more jets. In order to test the distinction between 2 versus 3 jets, events with more than 3 jets were excluded (about 10% of training events were removed). The binned significances obtained from the training with this boolean variable with and without removing events with $N_{jets(p_T>30)} > 3$ are 15.44 and 15.43, respectively. Hence, the power came from the distinction between 2 versus 3 jets small, and the variable $N_{jets(p_T>30)}$ was included in the input variable lists.

The distinctive signature of the two jets in the VBF production can be used to improve the BDT training performance for the VBF region, variables of the di-jet system are included in the input variable lists. The input variable optimization was performed separately for $e\tau_{had}$ and $\mu\tau_{had}$ channels, but the optimized variable lists were found to be very similar, a compromise between them is used as a result. Table 11.4 shows a summary of the selected variables in nonVBF and VBF regions. The shapes of each variable used in the BDT trainings in the VBF region, $\mu\tau_{had}$ channel, is shown in figure 11.8 and figure 11.9 for BDT₁ and BDT₂, respectively.

nonVBF	VBF
$m_{coll}(\ell, \tau_{had-vis})$	
$m_{vis}(\ell, \tau_{had-vis})$	
$m_T(\tau_{had-vis}, E_T^{miss})$	
$m_T(\ell, E_T^{miss})$	
E_T^{miss}	
$p_T(\ell)$	
$p_T(\tau_{had-vis})$	
$\eta(\tau_{had-vis})$	
$\Delta R(\ell, \tau_{had-vis})$	
$\Delta\eta(\ell, \tau_{had-vis})$	
$\delta\Delta\phi$	-
$\Delta\phi(\ell, \tau_{had-vis})$	-
-	$\Delta\phi(\ell, E_T^{miss})$
-	$N_{jets(p_T>30)}$
-	m_{jj}
-	$ \Delta\eta_{jj} $

Table 11.4: List of input variables used in the nonVBF and VBF regions. the selected variables are the same for the $e\tau_{had}$ and $\mu\tau_{had}$ channels.

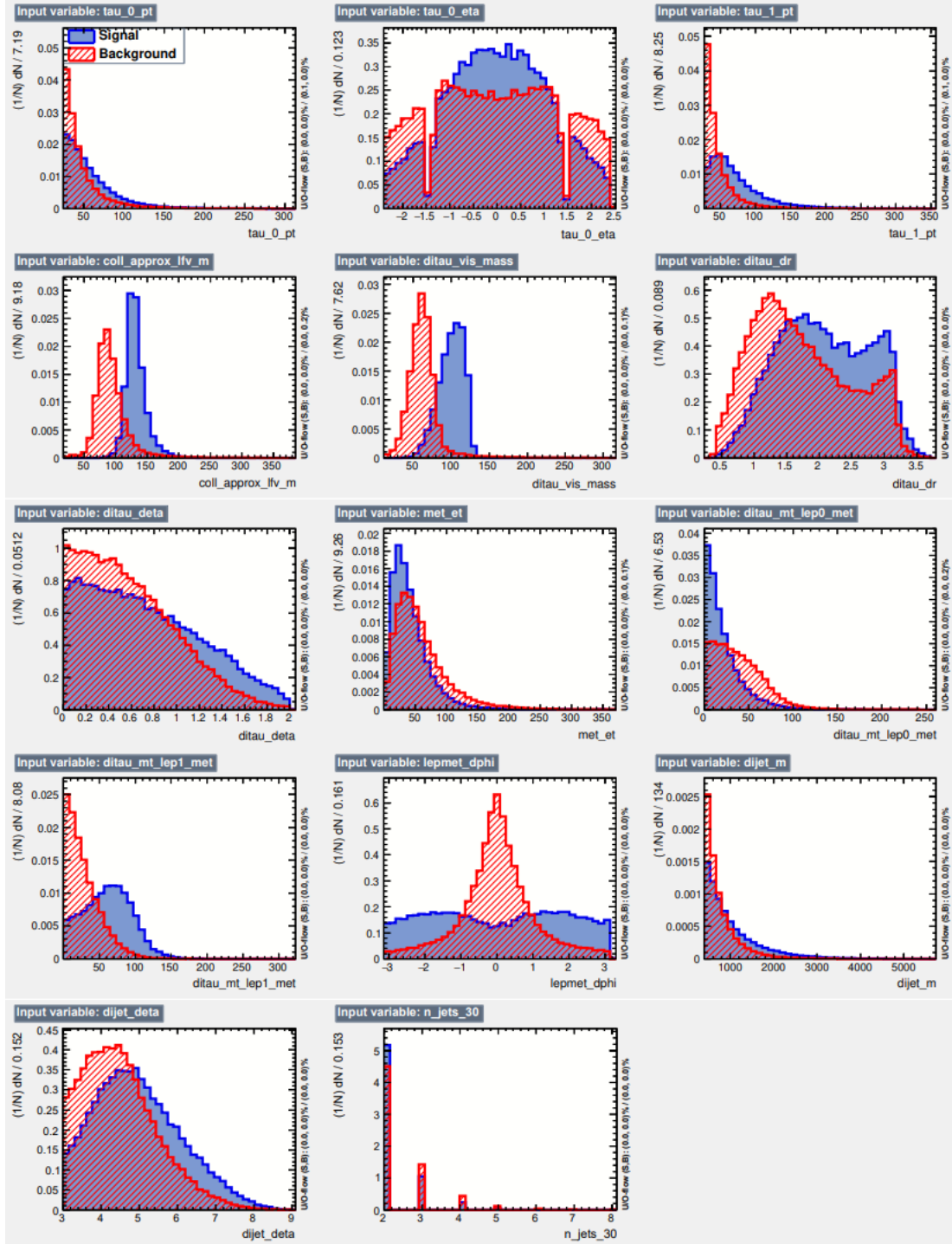


Figure 11.8: The shapes of the input variables used in the BDT₁ training in the VBF region, μ_{Thad} channel.

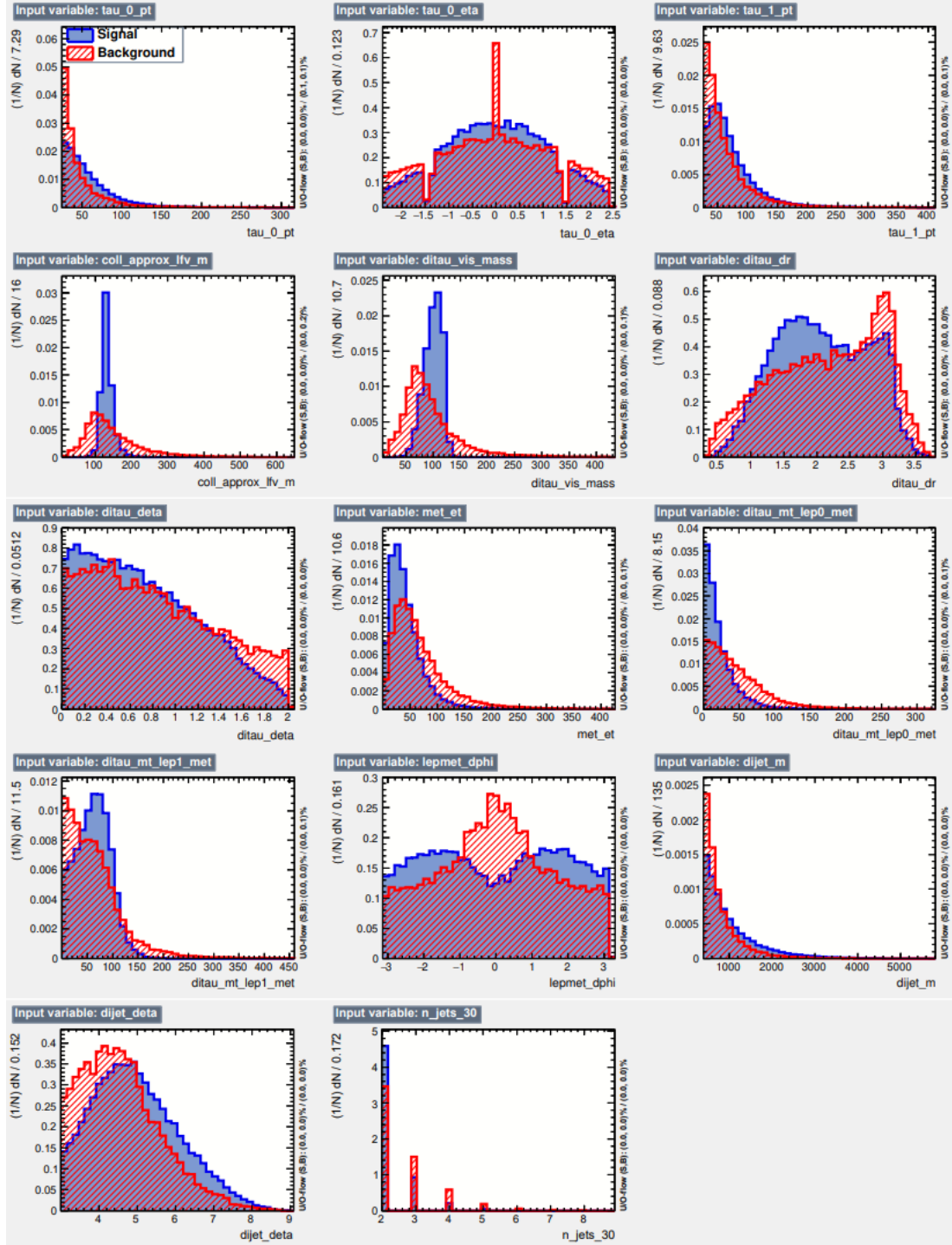


Figure 11.9: The shapes of the input variables used in the BDT₂ training in the VBF region, μ_{Thad} channel.

Chapter 12

Statistical and Systematic Uncertainties

The uncertainties of a measurement due to the fluctuation in data are presented as the statistical uncertainties. While the fluctuations of a measurement due to the imperfect knowledge involved in each step to get the final measurement are presented as the systematic uncertainties. Ideally, there is no fluctuation in the data sample if the detector has perfect response. However, it is impossible for the real experiments, all the high-energy physics experimental results should be presented with uncertainties. In the ATLAS experiments, the background and signal predictions are affected by different sources of uncertainties. The systematic uncertainties are considered in the statistical analysis as the nuisance parameters (NPs), as discussed in section 13.

In this chapter, the uncertainties arise from each step of the Higgs LFV analysis discussed in this thesis are introduced, with a focus on the MC-template method. In addition, an introduction to the uncertainties in the efficiency corrections used in the symmetric method is also presented. The experimental uncertainties arise from the detector and the reconstructed objects are discussed in section 12.2.1, the uncertainties in the misidentified background estimation are discussed in section 12.2.2, the uncertainties assigned to the $Z \rightarrow \ell\ell$ background normalization in section 12.2.3, the uncertainties arise from the efficiency correction procedure is discussed in section 12.2.4 and the theoretical uncertainties are discussed in section 12.2.5.

12.1 Statistical Uncertainties

In a measurement, the uncertainties on some randomly changed observations obtained by repeating the experiment is called statistical uncertainties. These uncertainties represent the statistical fluctuation of the sample due to the randomness of observations. There are two types of statistical uncertainties considered, which are:

- **Data Statistical Uncertainties**

The number of data events in each bin of a distribution is assumed to follow the Poisson distribution, with the expectation value being the observation within a certain bin. While the statistical uncertainties arise from the size of the dataset in each bin can be estimated as $\sqrt{N}$, where N is the number of data events in each bin. The statistical uncertainties reflects the fluctuation in data and regardless of the nuisance parameters.

- **MC statistical Uncertainties**

The statistics of the MC events used for template fits are limited, the statistical uncertainties of a measurement arise from the limitation. In the Higgs LFV analysis, the MC samples include the signal and background contributions (except the data-driven misidentified background). In the fit model for estimating the MC statistical uncertainties, each background sample are combined into a total background contribution, and each signal sample are also combined into one. The contribution from each signal (and background sample) in an individual bin of a distribution is represented as a weight value. The corresponding statistical uncertainties can be estimated as $\sqrt{\sum_{i=1}^n w_i^2}$, where n is the number of signal or background samples and w_i is the weight of the i^{th} signal or background sample.

12.2 Systematic Uncertainties

Systematic uncertainties arise from the limited knowledge of the detector and the analysis procedure, there is no measurement done with perfect knowledge and without any loss of information. The systematic uncertainties are used in the statistical analysis to extract the signal strength.

12.2.1 Experimental Uncertainties

The uncertainties originate from the detector (see section 4.2) and the reconstructed physics objects (see section 6) are grouped into the experimental uncertainties. Common experimental uncertainties related to the reconstructed objects could be coming from the resolution, energy scale, reconstruction, identification, isolation and trigger.

The experimental uncertainties are estimated following the recommendations provided by the combined performance groups, using the software tools provided with *AnalysisBase-21.2.119*. The shape of the score distributions, the number of background events and the signal cross-section could be affected by the experimental uncertainties. Each of the considered experimental uncertainties are listed and discussed in this section as follow.

12.2.1.1 Electron Uncertainties

The techniques used to estimate the electron uncertainties are described in [147]. The recommendations are provided by the ATLAS *ElectronGamma* group. The type of the considered electron uncertainties are listed as follow:

- Resolution: The energy resolution corrections uncertainties, all impacts are full correlated, there is only one single nuisance parameter from this source.
- Energy scale: The energy scale calibration uncertainties, similar to the energy resolution, only one nuisance parameter from this source.
- Reconstruction: The reconstruction efficiency scale factor uncertainties, which follows the *TOTAL* correlation scheme, and only one nuisance parameter combined all the respective impacts is used.
- Identification: The identification efficiency scale factor uncertainties, which follow the *SIMPLIFIED* correlation scheme. The full covariance matrix is split into 16 correlated sources of

systematic uncertainties and 12 uncorrelated statistical uncertainties, i.e. there are 28 nuisance parameters from this source.

- Isolation efficiency scale factor uncertainties: which follows the TOTAL correlation scheme, and only one nuisance parameter combined all the respective impacts is used.
- Charge identification: scale factor uncertainties, which corrects the charge misidentification rates in the MC samples to that measured in data.
- Trigger: trigger efficiency scale factor uncertainties, only one nuisance parameter from this source.

The TOTAL scheme is only used if the corresponding nuisance parameter is neither over-constrained nor pulled, as this scheme is sensitive and could be overestimating the uncertainty. The electron identification uncertainty was found to have large impact on the fit results, which are not consistent with the obtained impacts originate from the electron reconstruction uncertainty and isolation uncertainty. Therefore, the SIMPLIFIED scheme is used for the electron identification uncertainties.

12.2.1.2 Muon Uncertainties

The techniques used to estimate the muon-related uncertainties are described in [148]. The recommendations are provided by the *muon combined performance* group. The type of the considered uncertainties are listed as follow:

- Resolution: The momentum resolution uncertainties, derived from the variations in the resolution of the tracks from the inner detector and muon spectrometer. Two uncertainties were also derived from the variations in the charge-dependent momentum scale, based on the residual charge-dependent corrections between the data and MC samples. Hence, there are four nuisance parameters from this source.
- Energy scale: The muon uncertainties derived from varying the momentum scale. Only one nuisance parameter from this source.
- Reconstruction/Identification: The construction and identification efficiency scale factor uncertainties, include both statistical and systematic uncertainties. They are split into uncertainties for high-momentum ($p_T > 15$ GeV) and low-momentum ($p_T < 15$ GeV). Hence, there are four nuisance parameters from this source.
- Isolation: The isolation efficiency scale factor uncertainties, there are two nuisance parameters: the statistical and systematic uncertainties from this source.
- Trigger: The trigger scale factor uncertainties, there are two nuisance parameters: the statistical and systematic uncertainties from this source.

12.2.1.3 Tau Uncertainties

The tau related uncertainties are estimated based on the methods described in [149]. The recommendations are provided by the ATLAS *tau working group* with the recommended *2019-summer* set of uncertainties, which are described as follow:

- **Energy scale:** There are two in-situ energy calibration uncertainties, which are derived from a χ^2 based fit to data of the visible mass of the $\mu\tau$ pair in a dedicated tag and probe analysis. One detector uncertainty was derived from the energy or momentum measurement. Two additional uncertainties have been derived from the comparison of the reconstruction versus the distributions of the truth tau energy between the $Z \rightarrow \tau\tau$ Powheg+Pythia8 sample and samples with alternative geometries and GEANT4 physics-lists. Hence, there are five nuisance parameters from this source.
- **Reconstruction:** The tau reconstruction uncertainties which were derived by comparing the reconstruction efficiency of $Z \rightarrow \tau\tau$ Powheg+Pythia8 sample and samples with alternative geometries and GEANT4 physics-lists. Only one nuisance parameter from this source.
- **Identification:** There are ten tau RNN identification uncertainties, one refers to the tau identification efficiency scale factor uncertainties. The other eight uncertainties are derived separately for 1-prong and 3-prong, and for each bin with different p_T range (bins with p_T within 20-25, 25-30, 30-40 and greater than 40 GeV). These uncertainties have been derived by comparing the tau RNN identification efficiency in the data and MC samples in a dedicated $\mu - \tau$ tag and probe analysis. The last one have been derived by comparing different MC samples within the $p_T > 100$ GeV regime. Hence, there are ten nuisance parameters from this source.
- **eVeto on fake tau:** The uncertainties originate from the electron veto on the fake $\tau_{had-vis}$, which were derived by comparing the EBDT identification efficiency in the data and MC samples in a dedicated $e - \tau$ tag and probe analysis. Two nuisance parameters from this source.

12.2.1.4 Jet Uncertainties

The jet related uncertainties are estimated based on the methods described in [150]. The recommendations are provided by the ATLAS *JetEtMiss* group, following the recommended *R4 CategoryReduction FullJER* uncertainty set scheme, which are listed as follow:

- **Energy scale:** There are 30 jet-energy-scale (JES) uncertainties considered, which were derived for counting different impacts. These uncertainties include impacts from the pile-up effects, the η -intercalibration, flavour-tagging, the punch through effects¹, MC modelling, single-particle with high p_T , in-situ measurements and statistical uncertainty. There are 30 nuisance parameters from this source.
- **Energy resolution:** There are 13 jet-energy-resolution (JER) uncertainties considered, the detail of the derivations is described in [151]. There are 26 nuisance parameters from this source.
- **Flavour:** There are two flavour uncertainties considered, which originate from the effects of the jet-flavour composition and the flavour dependence of the jet response.
- **Jet vertex tagger:** There are two jet-vertex-tagger (JVT) efficiency scale factor uncertainties considered, one originate from the central region of the detector, and the other originate from the forward region.

¹Which refers to those jets passed through the calorimeter

- b-tagging: The b-tagging uncertainties are originated from the corrections of the b-tagging efficiency between the data and MC samples. The flavour-tagging uncertainties are decorrelated by using a eigenvector decomposition method, there are three nuisance parameters considered for the b-jets, another three nuisance parameters considered for the c-jets and four nuisance parameters considered for the light-flavour jets. Two additional nuisance parameters are considered for the extrapolation scheme in order to account the impacts from the high- p_T jets.

12.2.1.5 Missing Transverse Energy Uncertainties

The uncertainties on the missing transverse energy are estimated by using the methods described in [152]. The recommendations are provided by the JetEtMiss group. There are three uncertainties associated with the E_{miss}^T estimation, which are derived based on the soft term of the track-based E_{miss}^T for counting the impacts from the tracks which are not associated with the primary vertex. One nuisance parameter is used for the energy resolution corrections, the other two nuisance parameters are used for the energy scale calibration.

12.2.1.6 Luminosity Uncertainty

The uncertainty in the combined 2015-2018 integrated luminosity is 1.7% [153], obtained using the LUCID-2 detector [154] for the primary luminosity measurements.

12.2.2 Uncertainties in the Misidentified Background

The ABCD method used for the fake lepton estimation in the $\ell\tau_{\ell'}$ channels is described in section 10.1.1.1, the systematic uncertainties originate from the fake lepton estimation are derived following the guidelines of the Isolation Fake Forum. As mentioned, dedicated control regions are defined for estimating the fake lepton contribution in the signal region. The assigned systematic uncertainties have been derived for counting the impacts from the differences between the control regions and the signal regions. There are several sources these systematic uncertainties which are used in the statistical analysis, which are listed as follow:

- Statistical Uncertainty: The uncertainty arise from the limited statistics in the calculation of the transfer factor, which is considered independently for b-tag and b-veto events, as well as single lepton and dilepton triggered events.
- Flavour Composition: The OSF and SSF control regions are defined for the calculation of the transfer factor with data. The fake lepton contribution in the OSN signal regions is estimated by applying the transfer factor to the SSN background region, the corrections between the $SSN \times TF$ prediction and the OSN shape of the BDT is the one of the systematic uncertainty sources, which is originated from the composition difference between the CR and SR.
- Closure: The uncertainty assigned to the data-driven closure test which is performed by redefining the control regions. The fake lepton contribution in the signal region is re-estimated with the transfer factor which is recalculated in the redefined control regions. The closure uncertainty is assigned as the corrections between the data and the background prediction.
- Prompt lepton subtraction: The fake enriched control regions are defined for estimating the fake lepton contribution, and prompt leptons were subtracted in these CRs. An associated

uncertainty is assigned to the subtraction processes in the CRs. The subtracted MC backgrounds are different in each CRs according to the size of the systematic uncertainties.

- $d0$ bias: The $d0$ cut can change the fake composition. Another $d0$ cut is tested and compared with the nominal results, the difference between results with the new $d0$ cut and the nominal results is assigned as a systematic uncertainty.

The fake factor method is used for estimating the jet misidentified as $\tau_{had-vis}$ contribution in the $\ell\tau_{had}$ channels, the systematic uncertainties related to the fakes are originated from the calculation of the number of events with fake $\tau_{had-vis}$ in the signal region. From each step of the calculation discussed in section 10.1.2.1, the associated uncertainties are assigned to each individual term, such as the individual fake factor of each process, the R_{QCD} , the isolation factors, etc. All the assigned uncertainties for the fake τ_{had} estimation are described as follow:

- Individual fake factors: The statistical uncertainty on the individual fake-factors for the QCD multi-jets (F_{QCD}) and the W +jets (F_W) processes. These uncertainties arise from the limited statistics in the corresponding QCD control region and W +jets control region, which are defined for the fake-factors calculation. The fake-factors are calculated as a function of $p_T^{\tau_{had}}$ for 1-prong and 3-prong final states, the distributions are shown in figure 10.5, separate uncertainties are assigned to each p_T bin of these fake-factors distributions.
- R_{QCD} : The uncertainties associated with the R_{QCD} estimation, which includes the statistical uncertainty arise from the limited statistics in the regions where the R_{QCD} is calculated. The isolation factors in equation 10.9 are defined for the R_{QCD} estimation, statistical uncertainty arise from the limited statistics is also assigned to the regions where the isolation factors were calculated. In addition, the uncertainty arise from the difference between the isolation factors calculated using same-sign and opposite sign event selection is also considered.
- MC subtraction: In the fake-factors and R_{QCD} calculation, uncertainty arise from the subtraction of the MC samples is considered. This uncertainty has been evaluated by scaling the corresponding MC events by 10%, for example, the MC events are scaled by 1.1 (0.9) for the "up" ("down") variation. The 10% is a chosen conservative value considering the uncertainties on the cross-section MC subtracted processes.
- anti- $\tau_{had-vis}$ region: The statistical uncertainty arise from the limited statistics in the anti- $\tau_{had-vis}$ region.
- SS closure: Similar to the leplep channels, a non-closure systematic uncertainty is assigned to the closure test which is performed by repeating the fake-factor estimation but with the same-sign requirement, which means all the selections remain unchanged but require that the $\tau_{had-vis}$ and the lepton should carry the same sign of electric charge. In this way, the fake background contribution in the SS region is enhanced and the other background processes are suppressed. The BDT score distributions shown in figure 12.1 were generated within the SS regions, and slight discrepancies can be found between the data and background. The non-closure systematic uncertainty is assigned as the corrections between the data and background, which is derived with a SS ratio that can count discrepancies:

$$ratio(SS) = \frac{N_{Data}^{SR(SS)} - N_{MC}^{SR(SS)}}{N_{fakes}^{SR(SS)}} \quad (12.1)$$

Then multiply with the OS fake background contribution N_{fakes}^{SR} by this SS ratio. In order to reduce the impact of the statistical fluctuations in the SS ratios, a smoothing procedure based on the *ROOT* software was implemented before multiplying with the N_{fakes}^{SR} . Figure 12.2 the SS ratios before and after smoothing.

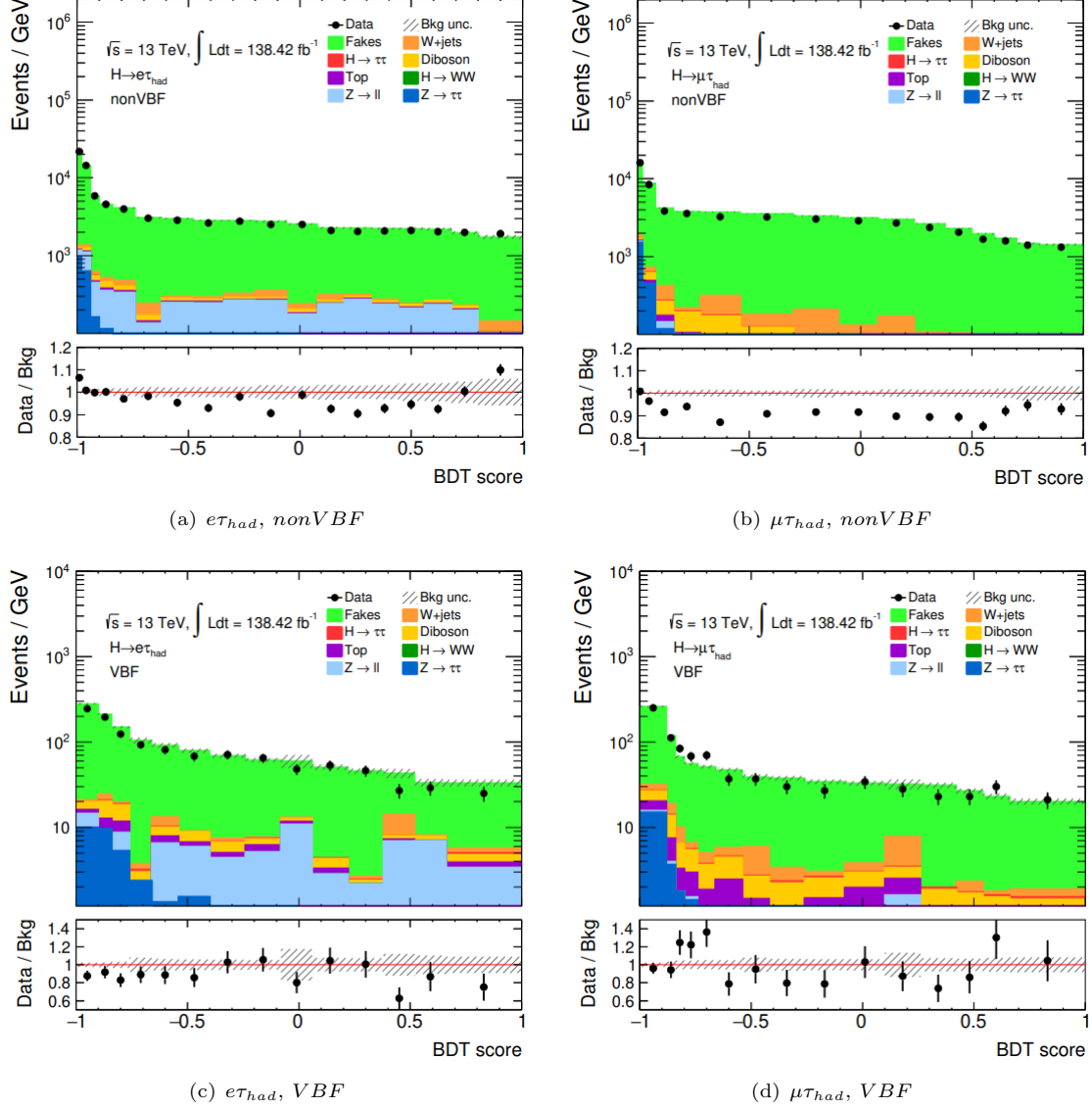


Figure 12.1: BDT score distributions in the SS nonVBF (top) and VBF (bottom) regions, for $e\tau_{had}$ (left) and $\mu\tau_{had}$ (right) channels.

12.2.3 $Z \rightarrow \ell\ell$ Background Normalization

In the $\ell\tau_{\ell}$ channels, a dedicated $Z \rightarrow \mu\mu$ control region is defined for studying this process, as discussed in section 10.1.1. The discrepancy between the data and MC samples is used for deriving a systematic uncertainty for the $Z \rightarrow \mu\mu$ background.

In the $\ell\tau_{had}$ channels, a dedicated $Z \rightarrow \ell\ell$ validation region is defined for validating the modelling of the $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ processes, as discussed in section 10.1.2. The discrepancy between the data and MC samples in this validation region is used as a systematic uncertainty.

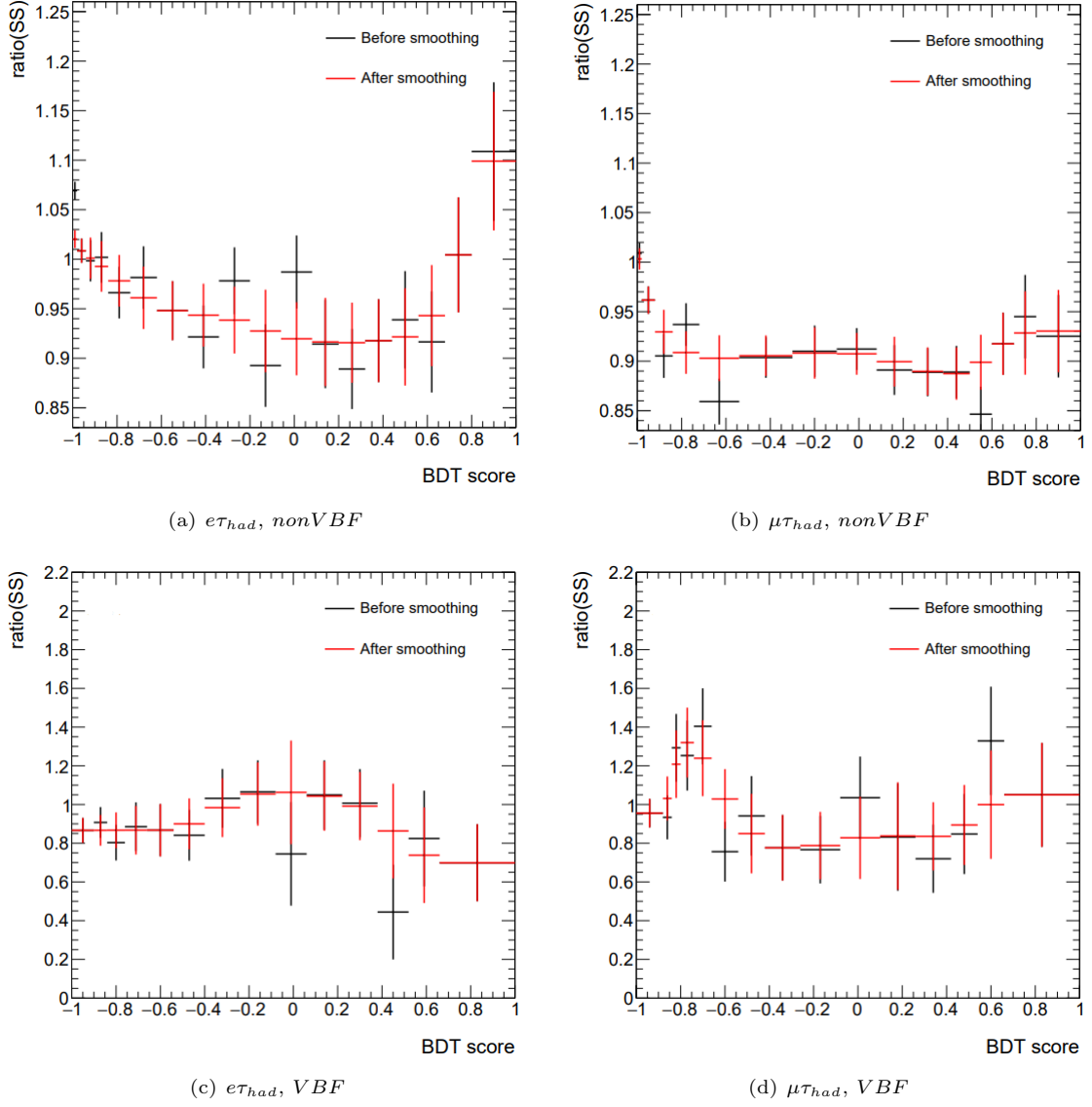


Figure 12.2: The SS ratios used to calculate the SS closure systematic uncertainty, for nonVBF (top) and VBF (bottom) regions, for $e\tau_{had}$ (left) and $\mu\tau_{had}$ (right) channels. The ratios before and after smoothing are both shown for comparison.

12.2.4 Uncertainties in the Efficiency Corrections

The uncertainties arise from the efficiency correction procedure (see chapter 9) for both $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channels are introduced. The efficiency correction consists of the trigger, reconstruction, identification and isolation efficiencies, which are measured for electron and muon separately. The real and mirror efficiencies in the $\ell\tau_{\ell'}$ channels are also measured. As the studies of the $\ell\tau_{had}$ channels are still ongoing, only the basic set of uncertainties are discussed in this section. The efficiency uncertainties are provided by the ATLAS ElectronGamma group and muon combined performance group.

As discussed in section 9.1.2, the muon efficiencies are calculated as the product of the reconstruction, identification and isolation efficiencies, in addition with the dilepton trigger efficiency for leplep channels or the single muon trigger efficiency for lephad channels. The corresponding

uncertainties are assigned: the muon trigger, reconstruction, and isolation efficiency uncertainties, as the joint reconstruction and identification efficiency maps provided by the ATLAS muon combined performance group was used, the identification efficiency uncertainties are covered by the reconstruction efficiency uncertainties. The procedures for estimating the muon efficiency uncertainties are the same as discussed in section 12.2.1.2, both statistical and systematic uncertainties are assigned to each of the muon efficiencies.

The muon efficiencies are provided by the ATLAS muon combined performance group, while the electron offline efficiencies are composed of the MC efficiencies multiplied by the scale factors which are provided by the ElectronGamma group. The uncertainties associated with the electron scale factors are described in section 12.2.1.1. In addition, the uncertainties of the *real* and *mirror* efficiencies are correlated, they are represented by one nuisance parameter, respectively. Both statistical and systematic uncertainties are assigned to the MC efficiencies.

12.2.5 Theory Uncertainties

The uncertainties originate from the background and signal predictions, the choice of the MC generators (see section 5.2) and the factorization or normalization scale of the MC generators are described in this section.

12.2.5.1 Background predictions

The theoretical uncertainties of the $Z \rightarrow \tau\tau$, $Z \rightarrow \ell\ell$ and Top background predictions are described in this section, while the considered fake-related background uncertainties are described in section 12.2.2.

The considered sources of the uncertainties for the $Z \rightarrow \tau\tau$ and $Z \rightarrow \ell\ell$ are:

- Parton Density Function (PDF) central value: The uncertainties assigned to the PDF central value which are evaluated considering the standard deviation of the 100NNPDF replicas event weights of NNPDF3.0nnlo PDF set used in SHERPA.
- renormalization (μ_R) and factorisation (μ_F) scales: The uncertainties evaluated using the event-weights provided by SHERPA, the scales are varied by 0.5 and 2.0 [155], then the observed differences are considered as systematic uncertainties.
- Matrix element matching (CKKW): The uncertainties assigned to the jet-to-parton matching, which are evaluated using truth-level parametrisation as a function of jet multiplicity and p_T of the Z boson. Similar to the μ_R and μ_F scales uncertainties estimation, the scales used for calculating the jet overlap between the Matrix Element (ME) and Parton Showering (PS) steps are varied with respect to the nominal (20 GeV) [156]. While the up and down variations are set to 30 GeV and 15 GeV, respectively. The systematic weights are computed in term of the up and down variations as $w = \frac{up - down}{up + down}$.
- Resummation scale (QSF): Similar to the CKKW procedure, the uncertainties associated with the scale used for resummation of soft gluon emission are evaluated by scaling the nominal value by 2 and 0.5 for the up and down variations, respectively.
- α_S : The uncertainties originate from the strong coupling strength α_S is estimated using the event weights provided by SHERPA.

- PDF alternative value: The uncertainties assigned to the PDF alternative value which are evaluated by comparing the predictions from the NNPDF3.0nnlo PDF set (nominal) with MMHT2014nnlo68cl [157] and CT14nnlo [158] PDF sets.

The nuisance parameters for the Top background are considered as normalization factors in the $\ell\tau_{\ell'}$ channels. While in the statistical fit without the $\ell\tau_{\ell'}$ channels and its dedicated Top CR - when performing the standalone fit of the $\ell\tau_{had}$ channels, the following theoretical uncertainties are considered:

- PDF: The PDF uncertainties are built by combining the available PDF sets along with the 30 eigenvector variations [157].
- Matrix element: Under the same parton shower model, the matrix element uncertainties are evaluated by comparing the nominal samples with alternative samples generated using different matrix-element generator [159].
- Parton shower: Under the same matrix-element generator, the parton shower uncertainties are evaluated by comparing the nominal samples with alternative samples generated using different parton shower model [160].
- Final state radiation: The uncertainties originate from the final state radiation are evaluated by using event-weights from the nominal sample reweighting [161].
- Initial state radiation: The uncertainties originate from the initial state radiation are evaluated by using event-weights from nominal sample reweighting [161].

12.2.5.2 Higgs Boson Background and Signal Predictions

The theoretical uncertainties originate from the SM Higgs decay processes share the same sources of those considered for the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals, under the assumption that the Higgs boson production and decay modes (except LFV) follow SM predictions. The Higgs boson production cross-section uncertainties are obtained from [162], which arise from the following three main sources:

- QCD scale: The uncertainties originate from the missing higher orders in perturbation theory, which also covers the uncertainty on jet bin migration.
- The uncertainty originate from the non-perturbative parts of the calculation, which include parton shower and hadronization.
- The uncertainty originate from the experimental input parameters such as the PDFs and the strong coupling strength α_S .

In the statistical analysis, the MVA outputs are used as the final discriminant for the fit model. The number of events per bin in the score distribution can be obtained in term of the total cross-section, while the uncertainty is estimated based on the variations in the number of events predicted per bin in the final score distribution, the number of events per bin n_{total} is:

$$n_{total} = \sigma_{total} \times \mathcal{L} \times A \times f_{bin} \quad (12.2)$$

where σ_{total} is the total cross-section, $\mathcal{L}$ is the luminosity, A is the acceptance of the total sample in a signal region and f_{bin} is the fraction of events in a certain bin of the score distribution. Therefore, the theoretical uncertainties for the signal process can be factorized based on this equation into the uncertainties on the σ_{total} , uncertainties on A and the uncertainties on the shape of the score distribution. While the total cross-section uncertainty $\Delta\sigma_{total}$ is correlated to the luminosity uncertainty $\Delta\mathcal{L}$, and the shape uncertainty is correlated to the acceptance uncertainty from a given source.

The QCD scale uncertainties for the gluon fusion process on the total cross-section are evaluated in a way accounts the total cross-section and acceptance uncertainties, therefore, the QCD scale uncertainties for gluon fusion use a different treatment for avoiding double counting the uncertainty on the total cross-section. Instead, there are 9 sources of uncertainties which are recommended by the *LHC cross-section working group* (LHCXSWG) are considered. Which includes four scale variations uncertainties which are derived following the methods in [163], two VBF topology uncertainties that are derived following the methods in [164], two p_T^H -shape uncertainties and one top mass dependence uncertainty.

The QCD scale uncertainties for the VBF, VH and $t\bar{t}H$ processes are evaluated by varying the renormalization μ_R and factorization μ_F scales by factors of 2 and 0.5 with respect to the central value.

12.3 Uncertainty Preprocessing

All the systematic uncertainties described above input to the fit model as nuisance parameters, there are several steps of preprocessing of those NPs before performing the statistical analysis. The preprocessing procedure is aim to find and suppress the statistical noise from the systematic uncertainties with small variations: as some samples have relatively small sizes, the respective upward and downward systematic variations are also small. Therefore, the corresponding shapes may be dominated by statistical noise. Such noise could be input to the fit model by the corresponding NPs with incorrect and unintentional variations.

In the Higgs LFV analysis, all statistical fits, fit model cross-checks, uncertainty breakdowns and post-fit figures are proceeded with the TRExFitter software package [165], additional details are described in chapter 13. The TRExFitter package provides several optional functions that can be used for the uncertainty preprocessing: symmetrization, smoothing and pruning for the NPs. They are described as follow:

- **Symmetrization**

The symmetrization procedure is applied to systematic uncertainties that have only an up or down variation. The existing variation is mirrored to create a variation in the opposite direction. For uncertainties related to electrons, muons, taus, and jets, both up and down variations are provided. The new variation is calculated by taking the mean value of the absolute values of the up and down variations, while keeping their signs. In cases where the bins of these uncertainties have variations in the same direction as the nominal value, the larger variation is selected and mirrored to create a variation in the opposite direction.

- **Smoothing**

To mitigate the impact of statistical fluctuations on systematic variations, a smoothing

procedure is implemented using the TRExFitter package. The default smoothing algorithm used is "MAXVARIATION" in TRExFitter. The smoothing procedure is applied before symmetrization for one-sided variations and after symmetrization for other cases. This helps to flatten out the fluctuations and improve the stability of the systematic variations.

- **Pruning**

In order to stabilize the fit, NPs are pruned if their impacts are small. The systematic uncertainty is pruned if it has less than 0.1% relative contribution compared to the signal sample yield in signal regions or the total yields in the control regions. Those systematic uncertainties have negligible impact on the final fit results, removing them can stabilize the fit and speed up the runs. This pruning procedure only applied when performing blinded fits, while the following conditions are set for pruning NPs in the unblinded fits:

- The normalization effect is negligible if both the up and down variation has less than 0.1% relative contribution compared to the nominal.
- The shape effect is negligible if the relative difference between the up and down shape only histogram with the nominal histogram is in less than 0.1% in all bins.
- The statistical uncertainty is negligible if the relative value in a bin is less than 1%.

Chapter 13

Statistical Analysis

In nature science, the language of theoretical physics is mathematics, while the language of experimental physics is statistics. The correspondence between theoretical predictions and experimental observations can be quantified by performing statistical analysis or by comparing the experimental data with the theoretical predictions. For example, the discovery of the Higgs boson in the ATLAS and CMS experiments.

In LHC experiments, an experimental observation refers to an event that consists of measurements of observable quantities, such as a collision of particles. Variables are defined based on the measurements obtained from different events. A statistical model is constructed by defining a function that, for a given set of parameter values including the signal strength and uncertainties described in Chapter 12, returns the probability density function (PDF) corresponding to the observed data. The statistical analysis aims to perform a statistical fit using the constructed model to extract the analysis results.

In this chapter, the strategies used to develop a statistical model for the Higgs LFV analysis are introduced in Section 13.1 and Section 13.2. The procedure for performing a statistical fit to measure the significance of a signal, which quantifies the deviation from the background-only hypothesis, is also discussed. The details of hypothesis testing are provided in Section 13.3. The Higgs LFV analysis utilizes the TRExFitter software package to perform the statistical analysis. This package allows for the construction of the statistical model in the *HistFactory* format [166]. The specific details of the statistical fits conducted for the Higgs LFV analysis are discussed in Section 13.4.

13.1 Inference and Probability Distributions

At the LHC, various quantities are measured and recorded for each collision. These measurements can be affected by the fluctuations in detector response, stemming from imperfect resolution and other sources of noise. The outcome of these measurements cannot be predicted with certainty due to inherent randomness, but probabilistic predictions can be made.

Different types of probabilities can arise in different processes. Some probabilities are associated with repeatable processes, such as the probability of a jet being misidentified as a hadronic τ -lepton. Interpretations of probability can be broadly categorized into two main categories: frequentist probability and Bayesian probability. Frequentist probability is applicable only to repeatable processes, while Bayesian probability can be applied to a wider range of processes [167]. Further details on these concepts are presented in the next section (Section 13.1.1).

13.1.1 Statistical Inference

Statistical inference is the process of using data analysis to make inferences about the properties of an underlying probability distribution. In this procedure, an estimate $\hat{\theta}$ of unknown parameters θ is derived from the observed dataset, and hypothesis testing is performed. The results of statistical inference can be presented as an interval estimate, which provides an estimate of the true value along with a probability at a certain confidence level. Additionally, the results can also be presented as the rejection of a hypothesis or the significance of a signal.

During the statistical inference procedure, the statistical uncertainties of the data sample, theoretical uncertainties, and systematic uncertainties (described in Section 12) should be considered and quantified in the final results. An important aspect of statistical inference is to distinguish the parameter of interest (POI) (which refers to the goal of a measurement) from nuisance parameters that are not directly of interest but must be accounted for in hypothesis testing. The following definitions are relevant:

- **Parameter of Interest**

The probability distribution of variables are obtained by the theoretical predictions. The theoretical models depend on unknown parameters, the parameter of interest (POI) refer to the goal of the final measurement. In the high-energy physics experiments, the POI mostly refer to the signal strength μ derived in the statistical analysis, additional details are provided in section 13.3.

- **Nuisance Parameters**

The detector response can be described by a probability model that depends on unknown parameters. The nuisance parameter θ refers to any parameters that are not the direct interest of the final measurement. For example, the modelling of the detector response are only the auxiliary measurements to obtain the precise experimental distributions. The nuisance parameters are determined together with the POI in the statistical inference procedure. In the Higgs LFV analysis, the nuisance parameters considered include the systematic uncertainties described in section 12 and the normalization factors.

Different statistical inference procedures have been established, such as the frequentist inference and Bayesian inference. However, they are not mutually exclusive. The descriptions of these two inferences are given as follow:

- **Frequentist Approach**

The interpretation of frequentist approach is based on the frequentist probability, which interprets the probability in equivalent term to "frequency". In other word, the relative frequency of an event occurred in an ensemble of repeatable experiments is defined as the probability of the event. The frequentist approach is used in most of the statistical analyses in the high-energy physics experiments, which is also the inference used in the Higgs LFV analysis. The results of frequentist procedure is a decision from hypothesis testing or a confidence interval, with the specified POI and test statistic (see section 13.3.2), the details of this approach are introduced in the remaining sections of this chapter.

- **Bayesian Approach**

Bayesian inference is based on the interpretation of Bayesian probability, where probability

measures one's degree of belief in an event, and this degree of belief can depend on previous measurements. Bayesian procedure uses Bayes' theorem to update the degree of belief based on the newly observed data. The main difference comparing to the frequentist approach is that the POI and NPs are fixed in frequentist inference rather than as being random variables. The results from the Bayesian approach could be a probability distribution rather than the hypothesis testing.

13.1.2 Discrete Probability Distributions

The function that associates the probability $P(x_i)$ of each of the possible values x_i of a random variable x is known as *probability distribution*. While the probability distribution of a discrete random variable is also known as *probability mass function*. In the case that there are n random variables, the *joint probability* $P(x_1, x_2, \dots, x_n)$ is the probability that these variables take the values given by the n -tuple $x_1, x_2, \dots, x_n$ simultaneously.

In statistical analysis, there are many types of probability distributions to describe the probabilities of the events, such as the most common probability distributions: the binomial distribution, Gaussian distribution and Poisson distribution, and they are described as follow:

- **Binomial Distribution**

The binomial distribution can be used to answer the typical question such as the probability of getting n successes in N independent Bernouli trials¹, assume that the probability of success is p , the probability mass function is given by:

$$P(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \quad (13.1)$$

- **Poisson Distribution**

The poisson distribution is a discrete probability distribution that expresses the probability of a given number of events within a fixed discrete interval (or a bin), under the assumption that these events occur with a known constant mean rate. Consider that N in equation 13.1 is a very large number, but p is very small such that the *rate parameter* $v = pN$ is finite. Then, if $N \rightarrow \infty$, equation 13.1 can be rewritten as the Poisson distribution in term of v :

$$P(n; v) = \frac{v^n e^{-v}}{n!} \quad (13.2)$$

The expectation number and the variance of a random variable that follows the Poisson distribution are both equal to v . In the Higgs LFV analysis, the score distributions are chosen as the final discriminant of the statistical analysis, each bin of the score distribution is assumed to follow the Poisson distribution. The distributions are used in the maximum likelihood estimate, as discussed in section 13.2.

- **Gaussian Distribution**

The Gaussian distribution or normal distribution is a continuous probability distribution. If the value of v in equation 13.2 is large, the Poisson distribution converges to a Gaussian distribution with average value v and standard deviation $\sqrt{v}$. The general form of Gaussian

¹Bernouli trials refer to experiments with boolean outcome and with fixed probability of success.

distribution is presented as probability density function (PDF), the details of PDF are given in the next section. The general form with mean value μ and variance σ is:

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \quad (13.3)$$

13.1.3 Probability Density Functions

The concept of a discrete random variable can be generalized to a continuous random variable, where the probability of an event is represented by a probability density function (PDF). For a continuous random variable, the probability of an event E within a certain interval is given by the integral of the PDF over that interval:

$$P(E) = \int_E f(x_1, x_2, \dots, x_n) d^n x \quad (13.4)$$

Assuming an event E is a subset of Ω $E \subseteq \Omega$, the normalization condition is also applied to the continuous case, the total area of the density function is $\int_{\Omega} f(x_1, x_2, \dots, x_n) d^n x = 1$.

13.2 Likelihood Function

One of the main aspects of statistical inference is determining the central value of a parameter based on observed data. An estimator is a function used to obtain this central value. In frequentist inference, the most common estimation method is the likelihood function, which constructs the joint probability distribution of all measurements from the observed data.

For a given set of unknown parameters (POI and NPs), the likelihood function, denoted as L , returns the corresponding probability density function (PDF) based on the observed data. If there are n random variables $x_1, x_2, \dots, x_n$ and given unknown parameters, the joint PDF of these random variables is given by the likelihood function as:

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) = \frac{dP(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m)}{dx_1 dx_2 \dots dx_n} \quad (13.5)$$

For each event with n random variables $\mathbf{x} = (x_1, x_2, \dots, x_n)$, the probability density function for the entire sample with N events $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ can be represented as the likelihood function. If these events are independent, the likelihood function can be written as the product of the PDFs corresponding to individual event:

$$L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \boldsymbol{\theta}) = \prod_{i=1}^N p(\mathbf{x}_i; \boldsymbol{\theta}) \quad (13.6)$$

In order to simplify the calculations, it is convenient to rewrite the likelihood function (equation 13.6) in logarithm form. This allows for computing the summation instead of the product:

$$-\ln L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \boldsymbol{\theta}) = -\sum_{i=1}^N \ln p(\mathbf{x}_i; \boldsymbol{\theta}) \quad (13.7)$$

13.2.1 Maximum Likelihood Method

The estimation of parameters can be obtained by maximizing the likelihood function, which means finding the set of parameters corresponding to the maximum value of the likelihood function. This method is known as Maximum Likelihood Estimation (MLE) and is used to estimate the parameters of a given distribution using observed data.

Maximum Likelihood Estimation is employed to determine the parameters that result in the highest joint probability for the observed data. For the unknown parameters in the parameter space $\theta \in \Theta$, the parameter values $\hat{\theta}$ can be obtained by maximizing the likelihood function. This is done by differentiating Equation 13.7:

$$\frac{\partial \ln L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \boldsymbol{\theta})}{\partial \theta_i} = 0, \quad \hat{\theta} = \arg \max_{\theta \in \Theta} L(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \boldsymbol{\theta}) \quad (13.8)$$

The process of obtaining the parameter set that corresponds to the maximum value of the likelihood function is known as the "best fit" or simply "fit," as the optimal parameters are determined by fitting the theoretical PDF model to the observed data sample.

13.2.2 Binned Maximum Likelihood Fits

The procedure considered above is corresponding to the unbinned maximum likelihood fit. For a single variable, it is preferable to subdivide the range of the variables into discrete interval, or bins of a *histogram*. In each bin, the data sample in a certain interval is fitted to the number of entries in that bin. In the *binned maximum likelihood fits*, the number of observed values used in the fit are reduced down to the number of bins, which significantly improves the run speed of fits. The impact on the precision of the fit due to the lost in information is not sufficient if the number of entries in each bin is large enough.

In the Higgs LFV analysis, the number of entries in each bin of a (BDT or NN) score distribution which used in the fit as a final discriminant is assumed to follow the Poisson distribution, the mean value of each bin can be estimated from the theoretical predictions and unknown parameters. The parameter of interest in this analysis is the signal strength μ , and according to equation 13.2, the rate parameter can be expressed as $\mu s_i + b_i$ for bin i , where s_i and b_i are the mean number of signal and background events in the bin i , which can be obtained by:

$$s_i = s_{total} \int_i f_s(x; \boldsymbol{\theta}_s) dx, \quad b_i = b_{total} \int_i f_b(x; \boldsymbol{\theta}_b) dx \quad (13.9)$$

where $f_s(x; \boldsymbol{\theta}_s)$ and $f_b(x; \boldsymbol{\theta}_b)$ are the PDFs of the variable x for signal and background events, respectively. While s_{total} and b_{total} are the total mean number of signal and background events, respectively. Hence, the binned likelihood function for a Poissonian fit can be constructed based on equation 13.2 and equation 13.6 as:

$$L = \prod_{i=1}^N p(n_i | \mu s_i + b_i) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \quad (13.10)$$

where n_i is the number of entries in the bin i .

In the symmetry method, the efficiency corrections must be taken into account, and the full background estimate is expressed as equation 10.11. Hence, equation 13.10 for the symmetry method and for the search of $H \rightarrow \mu\tau_e$ signal should be rewritten as:

$$L = \prod_{i=1}^N p(n_i^{\mu\tau} | b_i^{\mu\tau} + \mu s_i) = \prod_{i=1}^N p(n_i^{\mu\tau} | \tilde{n}_i^{e\tau} + f_{FF,i}^{\mu\tau} + f_{MC,i}^{\mu\tau} + \mu(s_i^{\mu\tau} - \mathcal{R}^e s_i^{e\tau})) \quad (13.11)$$

where $n_i^{\mu\tau}$ is the number of detected events in bin i of the final discriminant in the $\mu\tau$ sample used for the search of $H \rightarrow \mu\tau$, the background prediction $b^{\mu\tau}$ in the sample includes the symmetric background $\tilde{n}^{e\tau}$ and the fake background contributions based on the fake-factor method and MC simulation. For the search of the $H \rightarrow e\tau$ signal, the equivalent equation is applied when swapping e and μ .

In both equation 13.10 and equation 13.11, the branching ratio of Higgs LFV decays is assumed to be 1%, while the amplitude of the signal in the observed data is represented with the signal strength as $\mathcal{B}(H \rightarrow \ell\tau) = \mu \times 1\%$.

The systematic uncertainties of the signal and background (see section 12) are included in the fits as a set of nuisance parameter θ , they are considered as additional Gaussian constraint terms in the likelihood function. In addition, the statistical uncertainties of the MC events in each bin are also considered via additional Poisson constraint terms in the likelihood function. However, there is no guarantee for an exact coverage of the uncertainty interval. To properly represent the uncertainty, a confidence interval [168] should be presented along with the estimate. The confidence interval demonstrates the range of variation around the estimated mean value and indicates the probability, or confidence level, that the true value falls within a certain interval. In the case of a Gaussian probability density function (PDF) model, the commonly used confidence levels are 68.27% (1σ) or 95.45% (2σ).

13.2.3 Simultaneous Fits

Instead of performing the fit with one POI for one channel, another option could be constructing the likelihood function for two different data samples from two channels which depends on a parameter set. In a *Simultaneous fit*, each channel has an associated Poisson model, and the likelihood function is the product of all the Poisson terms.

In the Higgs LFV analysis, two signals from the channels $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ are considered. A simultaneous fit for these two signals is performed using the MC-template method, the two POIs corresponding to $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ are estimated in the simultaneous fit of the $e\tau_\mu$, $e\tau_{had}$, $\mu\tau_e$ and $\mu\tau_{had}$ final states. In the 2-POI fits, the assumption applied in the 1-POI fit that the branching ratio of one signal is zero when fitting the other signal, can be removed. More statistical power is allowed to further constraint the nuisance parameters in the 2-POI fits. The details of the fitting strategy and the fit setup used in the each category and each method is described in section 13.4.

13.3 Hypothesis Testing

The hypothesis tests are used for discriminating which hypotheses that the observed data is supporting. Each event is assigned to one of the hypotheses depending on the predicted probability distributions under either of the considered hypotheses. In most tests, there are two hypotheses considered: the *background-only hypothesis* or *null hypothesis* H_0 , and the *alternative hypothesis* H_1 .

The signal significance can be obtained by quantifying the incompatibility between the data and the background-only hypothesis.

The probability model for the background-only hypothesis is constructed similar to the Poisson term in equation 13.10, but under the assumption that the rate parameter is the number of the background events b_i in bin i and $\mu = 0$: $p_{H_0}(n_i|b_i)$. In the alternative hypothesis, a clear evidence for the presence of a signal is an excess of events such that $\mu = 1$: $p_{H_1}(n_i|s_i + b_i)$. The hypothesis test is to check whether the observed data is more consistent with the background-only hypothesis rather than the alternative hypothesis.

The test statistics based on the profile likelihood ratio are defined in section 13.3.1. The sensitivity of an experiment can be characterized by the *statistical significance*, which can be obtained by the mean of a *p-value*, as introduced in section 13.3.2.1. If there is no discovery observed in the data or if the background-only hypothesis is true, the exclusion upper limits are set, the definition is introduced in section 13.3.2.2. In the estimation of the significance, the *Asimov dataset* is used, the details are described in section 13.3.3. The full descriptions of the test statistics and the asymptotic distributions defined for the hypothesis tests are given in [169].

13.3.1 Profile Likelihood

The histogram of the variables constructed from the specific control regions are used to constrain the nuisance parameters. For the histogram of one variable, the mean value or the rate parameter m_i of the Poisson distribution in bin i could be expressed in term of a quantity $u_i(\boldsymbol{\theta})$ that depends on the nuisance parameters $\boldsymbol{\theta}$. The background normalization parameters, the signal and background shape parameters are constrained by including this quantity in the likelihood function in equation 13.10 as:

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \prod_{j=1}^M \frac{u_j^{m_j}}{m_j!} e^{-u_j} \quad (13.12)$$

where M is the number of bins and j represents the j^{th} bin in the histogram of the variable constructed from the control region.

In order to test a hypothesized value of μ , a test statistic is constructed by taking the ratio of the maximized likelihood function in equation 13.12 with a certain condition, which is known as the *profile likelihood ratio*:

$$\lambda(\mu) = \frac{L(\mu, \hat{\boldsymbol{\theta}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad (13.13)$$

where $\hat{\boldsymbol{\theta}}$ represents the parameter set that maximizes the likelihood function with a specified value of μ . Therefore, the numerator is a conditional maximized likelihood function. While in the denominator, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the estimators of the maximized unconditional likelihood function. The values of the profile likelihood ratio are within the range $0 \leq \lambda \leq 1$, the higher value of λ represents that data is more compatible with the hypothesized value of μ .

13.3.2 Test Statistic

The profile likelihood ratio in equation 13.13 reduces the information of a dataset to only one value that can be used to perform the hypothesis test, such quantity is known as the *test statistic*.

A hypothesis test is typically performed with test statistic which has discrimination power between the background-only hypothesis and alternative hypothesis.

In general, the hypothesis test is performed based on the test statistic (equation 13.13) which is defined in term of logarithm:

$$t_\mu = -2\ln \lambda(\mu) \quad (13.14)$$

The higher value of t_μ represents the better agreement between the data and the μ value, while the applications of this form are described in the remaining part of this section.

The test statistic t_μ is also used to compute the *p-value* for quantifying the level of disagreement between the data and the hypothesis. The p-value is constructed with the PDF of t_μ and the observed value of the $t_{\mu, obs}$ as:

$$p_\mu = \int_{t_{\mu, obs}}^{\infty} f(t_\mu|\mu) dt_\mu \quad (13.15)$$

The p-value represents the probability p_μ of obtaining the test results which are more extreme than the observed results, under the assumption that the background-only hypothesis is correct. Therefore, a lower p-value represents the stronger disagreement between the observed data and the background-only hypothesis.

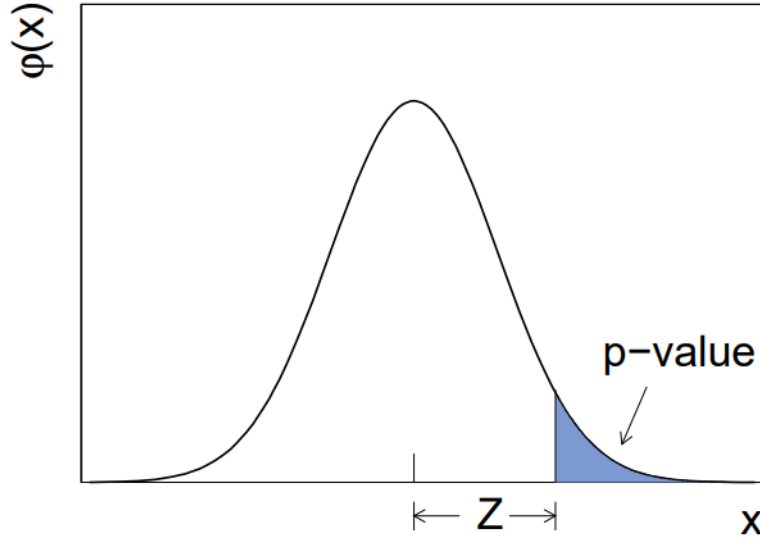


Figure 13.1: The relation between the p-value and significance Z under a normal distribution $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ [169].

In high-energy physics, scientists prefer to report the *significance level* of their results instead of the p-value. The significance level is reported as $Z\sigma$ for a corresponding obtained p-value, figure 13.1 shows the relation between Z and the p-value. Under a normal distribution $\varphi(x)$, the significance Z can be obtained in the integral:

$$p = \int_Z^{\infty} \varphi(x) dx = 1 - \Phi(Z) \quad (13.16)$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\Phi(Z)$ is the integration results in term of Z . In general, scientists claim an *evidence* of signal if the significance level is at least 3σ , which corresponds to a p-value of 1.35×10^{-3} . To claim a *discovery* of signal, the significance level should be at least 5σ , which corresponds to a p-value of 2.87×10^{-7} .

13.3.2.1 Test Statistic for Discovery

The value of the estimator $\hat{\mu}$ that maximized the likelihood function is not need to be non-negative. However, frequently it is required that the value of μ should be non-negative in a signal process. If one has estimated that $\mu \geq 0$ and $\hat{\mu} < 0$ in a hypothesis test, it also implies that the value of $\mu = 0$ refers to the highest consistency between the data and the value of μ . Hence, a profile likelihood ratio $\tilde{\lambda}$ can be defined separately in term of $\hat{\mu}$ with different values and the conditional maximum likelihood estimators $\hat{\theta}$ with different specified μ values:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0. \end{cases} \quad (13.17)$$

While the test statistic in equation 13.14 could also be defined separately and accordingly as $\tilde{t}_\mu$:

$$\tilde{t}_\mu = -2\ln \tilde{\lambda}(\mu) = \begin{cases} -2\ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ -2\ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0. \end{cases} \quad (13.18)$$

The presence of a positive signal could be represented as the rejection of the background-only hypothesis with $\mu = 0$. In this case, the test statistic for a positive signal could be defined according to equation 13.18: for the hypothesized value $\mu = 0$, if the maximum likelihood estimator $\hat{\mu} \geq 0$, then test statistic q_0 is $-2\ln \lambda(0)$, otherwise ($\hat{\mu} < 0$), $q_0 = 0$:

$$q_0 = \begin{cases} -2\ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0. \end{cases} \quad (13.19)$$

For a downward fluctuation in the data, the best fit estimator $\hat{\mu}$ is found to be negative. While for an upward fluctuation, an excess in data is presented, and the higher value of $\hat{\mu}$ give larger value of q_0 , which also means the stronger disagreement between the data and the background-only hypothesis.

In particle physics experiments, rejecting the background-only hypothesis is only part of a discovery of new signal, one would also need to declare the level of disagreement between the data and the hypothesis with $\mu = 0$ [169]. As introduced before, this could be presented by the p-value and signal significance Z , and in term of q_0 , equation 13.15 can be written as:

$$p_0 = \int_{q_0, obs}^{\infty} f(q_0|0) dq_0 \quad (13.20)$$

where $f(q_0|0)$ represents the PDF of q_0 constructed under the assumption of background-only

hypothesis. In order to test how the data is differed from the value $\mu' = 1$ being tested, the PDF of $f(q_0|\mu')$ is needed to compute the median of q_0 , and hence the median discovery significance.

13.3.2.2 Test Statistic for Upper Limits

In particle physics, if there is no significant signal presented in the data, an upper limit is set on the signal strength μ with certain confidence level by using the CL_s method [170]. Consider the signal process with $\mu \geq 0$, the test statistic $\tilde{q}_\mu$ for setting the upper limit can be defined according to equation 13.18. However, regarding the data with $\hat{\mu} > \mu$ will spoil the performance of the test for setting an upper limit. In this case, the test statistic $\tilde{q}_\mu$ is not considered in the hypothesis rejection and defined to be zero:

$$\tilde{q}_\mu = \begin{cases} -2\ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0, \\ -2\ln \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \quad (13.21)$$

The higher value of $\tilde{q}_\mu$ implies the stronger disagreement between the data and the hypothesized value of μ . The difference between the test statistic q_0 and $\tilde{q}_\mu$ is that when both of them equal to zero, $q_0 = 0$ refers to the downward fluctuation in the data and $\tilde{q}_\mu = 0$ refers to the upward fluctuation in the data. Which indicates the fact that q_0 is not the special test statistic with respect to q_μ with $\mu = 0$.

For the test statistic of $\tilde{q}_\mu$, the calculation of the p-value for representing the level of agreement is based on equation 13.15:

$$p_\mu = \int_{\tilde{q}_\mu, obs}^{\infty} f(\tilde{q}_\mu|\mu') dq_\mu \quad (13.22)$$

where $f(\tilde{q}_\mu|\mu')$ is the PDF of $\tilde{q}_\mu$ under the assumption of a hypothesis with non-zero value of μ' for setting an upper limit. In order to quote the upper limit with 95% confidence level, the PDF of $f(\tilde{q}_\mu|0)$ is needed to compute the p-value which is equal to 0.05.

13.3.3 Asimov Dataset

In order to quantify the sensitivity of a signal discovery, one needs to estimate the *median significance* with a non-zero value μ' . In other word, the PDF of $f(q_0|0)$ can be used to derive the p-value for the level of disagreement between data and the background-only hypothesis, while the PDF of $f(q_\mu|\mu')$ can be used to quantify the sensitivity under the alternative hypothesis. To obtain the $med[q_0|\mu']$, one need to derive the variance of the maximum likelihood estimator $\hat{\mu}$ of equation 13.14, which is assumed to follow the Gaussian distribution with mean value μ' . According to [169], the variance $\sigma = var[\hat{\mu}]$ can be derived with an artificial dataset which is known as *Asimov dataset*.

The Asimov dataset is constructed by suppressing all the statistical fluctuations in the dataset, and under the assumption that the number of events in each bin (n_i and m_j in equation 13.12) are equal to their expectation values such that [169]:

$$n_{i,A} = E[n_i] = \mu' s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}), \quad m_{j,A} = E[m_j] = u_j(\boldsymbol{\theta}) \quad (13.23)$$

where A refers to Asimov, s_i and b_i are the median value of the signal and background events in bin i . Hence, by performing the statistical analysis described above with the Asimov dataset, one can obtain the median significance.

Using Asimov dataset with $\mu' = 1$ to perform the hypothesis test and construct the corresponding test statistic $q_{0,A}$ as described in section 13.3.2.1, one can estimate the median or expected significance by following the calculations described in [169]:

$$\text{med}[Z_0|\mu'] = \sqrt{q_{0,A}} \quad (13.24)$$

While using the Asimov dataset with $\mu' = 0$ and constructing the corresponding test statistic $\tilde{q}_{\mu,A}$ as described in section 13.3.2.2, the expected upper limits at 95% confidence level can be obtained by following the procedure described above with the median p-value:

$$\text{med}[p_\mu|0] = 1 - \Phi(\sqrt{\tilde{q}_{\mu,A}}) \quad (13.25)$$

13.4 Fits Setup for the Higgs LFV Analysis

The Higgs LFV analysis exploited the TRExFitter [165] software package to create *RooFit* workspace [171] and proceed the statistical analysis. The TRExFitter is a framework bases on asymptotic formulae described in [169] to perform the binned profile likelihood fits, while the fit models are built in HistFactory format. The classes of the discovery significances, interval estimation and combined measurements are provided the *RooStats* project [172], which is built on top of the RooFit.

The statistical analysis is conducted through binned profile likelihood fits, where the likelihood function $L(\mu, \theta)$ is constructed as a product of Poisson probability terms over all bins of the final discriminant distributions. The BDT (or NN) score distributions are used as the final discriminant. Additionally, histograms with a single bin containing the number of events in the Top and $Z \rightarrow \tau\tau$ control regions are utilized to constrain the corresponding background event yields. The parameters of interest are $\mu_{e\tau}$ and $\mu_{\mu\tau}$, representing the branching ratios $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$, respectively. The nuisance parameters include the systematic uncertainties described in chapter 12 and the normalization factors. Systematic uncertainties are implemented as Gaussian constraints, while normalization factors are implemented as log-normal constraints in the likelihood function. The normalization factors are treated as free floating parameters in the fit.

In the MC-template method, the likelihood function with one parameter of interest (POI) is constructed considering various regions: the signal regions in each channel and category, the $Z \rightarrow \tau\tau$ control regions in each category and $\ell\tau_{\ell'}$ channels, and the Top control regions for each category and $\ell\tau_{\ell'}$ channels. Simultaneous 2-POI fits are performed for each category and channel in the signal and control regions. On the other hand, for the symmetry method, independent 1-POI fits are performed for each category and channel, as the results of the symmetry method are presented as branching ratio differences.

13.4.1 Independent 1-POI Fits

When performing the independent 1-POI fit for the $H \rightarrow e\tau$ search, the POI $\mu_{e\tau}$ is estimated by combining the $e\tau_\mu$ and $e\tau_{had}$ final states, while the branching ratio $\mathcal{B}(H \rightarrow \mu\tau)$ is set to zero.

Similarly, for the independent search of $H \rightarrow \mu\tau$ signal, the POI in the fit $\mu_{\mu\tau}$ is estimated by combining the $\mu\tau_e$ and $\mu\tau_{had}$ final states, and the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ is set to zero.

In each category, the fit results for the $\ell\tau_{\ell'}$ final states, combined with the fit results for the $\ell\tau_{had}$ final states, are chosen from the fits obtained using the MC-template method and symmetry method based on their performance. All the fit results are summarized in chapter 14. The MC-template method performs better than the symmetry method in the non-VBF category, while the symmetry method performs better than the MC-template method in the VBF category. Therefore, for the fits in the non-VBF category, the $\ell\tau_{had}$ channel is combined with the $\ell\tau_{\ell'}$ channel from the MC-template method. In the VBF category, the $\ell\tau_{had}$ channel is combined with the $\ell\tau_{\ell'}$ channel from the symmetry method. However, the studies of the $\ell\tau_{had}$ channel from the symmetry method are still ongoing, and they could potentially be combined with the other channels to obtain the full analysis results in the future.

In the MC-template method, as discussed in section 10.1.1 and section 10.1.2, the $Z \rightarrow \tau\tau$ normalization factors are calculated for both the $\ell\tau_{\ell'}$ channel and the $\ell\tau_{had}$ channel. Therefore, three $Z \rightarrow \tau\tau$ normalization factors are included in the fits as nuisance parameters: one from the non-VBF category of the $\ell\tau_{\ell'}$ channel, and two from the $\ell\tau_{had}$ channel. The Top normalization factors are calculated in the dedicated Top control regions in the $\ell\tau_{\ell'}$ channel. In the non-VBF category, the Top normalization factor is shared with the $\ell\tau_{had}$ channel. For the symmetry method, since no Top control region in the $\ell\tau_{had}$ channel is defined to constrain the normalization of the Top-quark background, uncertainties are instead assigned to the Top-quark cross-section.

Method	Channel	Category	Region	1-POI Fit	2-POI Fit
MC-template	$\ell\tau_{\ell'}$	nonVBF	SR	✓	✓
			$Z \rightarrow \tau\tau$ CR	✓	✓
			Top-quark CR	✓	✓
		VBF	SR		✓
			$Z \rightarrow \tau\tau$ CR		✓
			Top-quark CR		✓
MC-template	$\ell\tau_{had}$	nonVBF	SR	✓	✓
		VBF	SR	✓	✓
Symmetry	$\ell\tau_{\ell'}$	nonVBF	SR		
		VBF	SR	✓	

Table 13.1: The fit setup for the statistical analysis, the 1-POI and 2-POI fits considered in the combinations of channels, regions and categories are marked.

13.4.2 Simultaneous 2-POI Fits

The simultaneous 2-POI fits are performed by using the MC-template method, the parameters of interest are the branching ratios of both search: $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$, which are estimated simultaneously for the $e\tau_{\mu}$, $e\tau_{had}$, $\mu\tau_e$ and $\mu\tau_{had}$ final states. The assumption in the 1-POI fits that the branching ratio one search is set to zero in the fit of the other search is removed in the

simultaneous 2-POI fits.

In the simultaneous measurement, the $Z \rightarrow \tau\tau$ and Top normalization factors considered in the MC-template method are correlated between the SRs and CRs, but they are decorrelated between each category. As the $Z \rightarrow \tau\tau$ normalization factors are calculated independently for both $\ell\tau_{\ell'}$ channel and $\ell\tau_{had}$ channel, the normalization factors are decorrelated between both final states.

The statistical fit setup for the Higgs LFV analysis is summarised in table 13.1, the considered SRs and CRs in both method are listed, and the 1-POI and 2-POI fits performed in each category and region are specified.

Chapter 14

Run-2 Analysis Results

The results of the searches for lepton-flavour violating decays of the Higgs boson, obtained by performing the statistical analysis of the Run-2 dataset, are summarised in this chapter. The fit models used are described in chapter 13, both 1-POI and simultaneous 2-POI fits are performed. The results of the 1-POI fit of the $H \rightarrow e\tau$ (section 14.1.1) and $H \rightarrow \mu\tau$ (section 14.1.2) searches, and the simultaneous measurement of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals (section 14.2) are summarised in the following sections.

For the results described in each section, the best-fit μ values (POI) are presented, and the impact of different systematic sources on each POI are summarised. The pulls and post-fit constraints on the nuisance parameters are also discussed. The results for the Run-2 data were found to have a small excess but below the threshold of evidence of a new signal, the observed significances and upper limits are presented. The MVA technique described in chapter 11 is exploited to enhance the signal sensitivity, and the post-fit BDT score distributions obtained using the MC-template method are also shown. The results of the symmetry method are presented in terms of the branching ratio difference, which are measured in each category and presented together with the combined results in section 14.3.

14.1 Independent Searches for $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$

The results of the 1-POI fits performed for each search separately by combining the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ final states are summarised in this section.

14.1.1 1-POI Fits of $H \rightarrow e\tau$

When performing the 1-POI fit of the $H \rightarrow e\tau$ search, the branching ratio $\mathcal{B}(H \rightarrow \mu\tau)$ is set to zero. The standalone per-channel and per-category fits of the $H \rightarrow e\tau$ search were performed, while the combined fits for the combination of each category and each channel were also studied. For the standalone fits and the combined fits of the $H \rightarrow e\tau$ search using the MC-template method, the significance and the upper limits at 95% CL on the $\mathcal{B}(H \rightarrow e\tau)$ and the best-fit μ values are shown in table 14.1.

The $e\tau_{had}$ standalone fits are performed without the Top normalization factors, the Top theory uncertainties are included instead. The systematic uncertainties described in chapter 12 are grouped into several sources, the impact of those systematic sources on the POI of the combined $e\tau$ fit

(nonVBF+VBF) are listed in table 14.2, while the ranking of the top 20 systematic uncertainties which have highest post-fit impact on the POI is shown in figure 14.2. According to table 14.2 and figure 14.2, the fake background uncertainties are found to have the largest constraints. The post-fit constraints of the NPs are shown in figure 14.1, while the correlation matrix of the NPs is visualized in figure 14.3.

Results of the Fit with One POI: $H \rightarrow e\tau$				
Region	Category	Significance	Upper limit	Best-fit mu value
$e\tau_\mu + e\tau_{had}$	nonVBF+VBF	1.8987 (11.36)	$0.2099 (0.1172^{+0.0467}_{-0.0328})$	$0.1108^{+0.0592}_{-0.0585}$
$e\tau_\mu$	nonVBF+VBF	1.9835 (9.11)	$0.2991 (0.1657^{+0.0668}_{-0.0463})$	$0.1614^{+0.0822}_{-0.0810}$
$e\tau_\mu$	nonVBF	2.6975 (8.06)	$0.4131 (0.1866^{+0.0758}_{-0.0521})$	$0.2509^{+0.0970}_{-0.0936}$
$e\tau_\mu$	VBF	-1.3291 (4.06)	$0.2782 (0.4280^{+0.1713}_{-0.1196})$	$-0.2871^{+0.2163}_{-0.2222}$
$e\tau_{had}$	nonVBF+VBF	0.5437 (8.10)	$0.2113 (0.1761^{+0.0698}_{-0.0492})$	$0.0484^{+0.0894}_{-0.0892}$
$e\tau_{had}$	nonVBF	0.2792 (6.28)	$0.2672 (0.2437^{+0.0969}_{-0.0681})$	$0.0334^{+0.1235}_{-0.1244}$
$e\tau_{had}$	VBF	0.5178 (5.99)	$0.3006 (0.2527^{+0.1007}_{-0.0706})$	$0.0657^{+0.1277}_{-0.1269}$

Table 14.1: The standalone and combined 1-POI fit results of the $H \rightarrow e\tau$ search using the MC-template method, the observed (expected) significance, observed (expected) upper limits at 95% CL and the best-fit mu values are shown.

Source of Uncertainty	Impact on POI
Full Unc	$+0.059$ -0.058
Full Syst	$+0.051$ -0.051
MC stat	$+0.036$ -0.037
B-jet	$+0.007$ -0.007
Fake estimation $e\tau_{had}$	$+0.021$ -0.022
Fake estimation $e\tau_\mu$	$+0.028$ -0.027
Lepton	$+0.003$ -0.002
Luminosity	$+0.007$ -0.005
Jet + MET	$+0.011$ -0.011
Tau	$+0.008$ -0.007
Signal Theory Unc.	$+0.009$ -0.005
Zll normalization	$+0.001$ -0.001
Z Theory unc.	$+0.009$ -0.009

Table 14.2: Grouped impact of different systematic sources on the POI of the combined $e\tau$ fit (nonVBF+VBF).

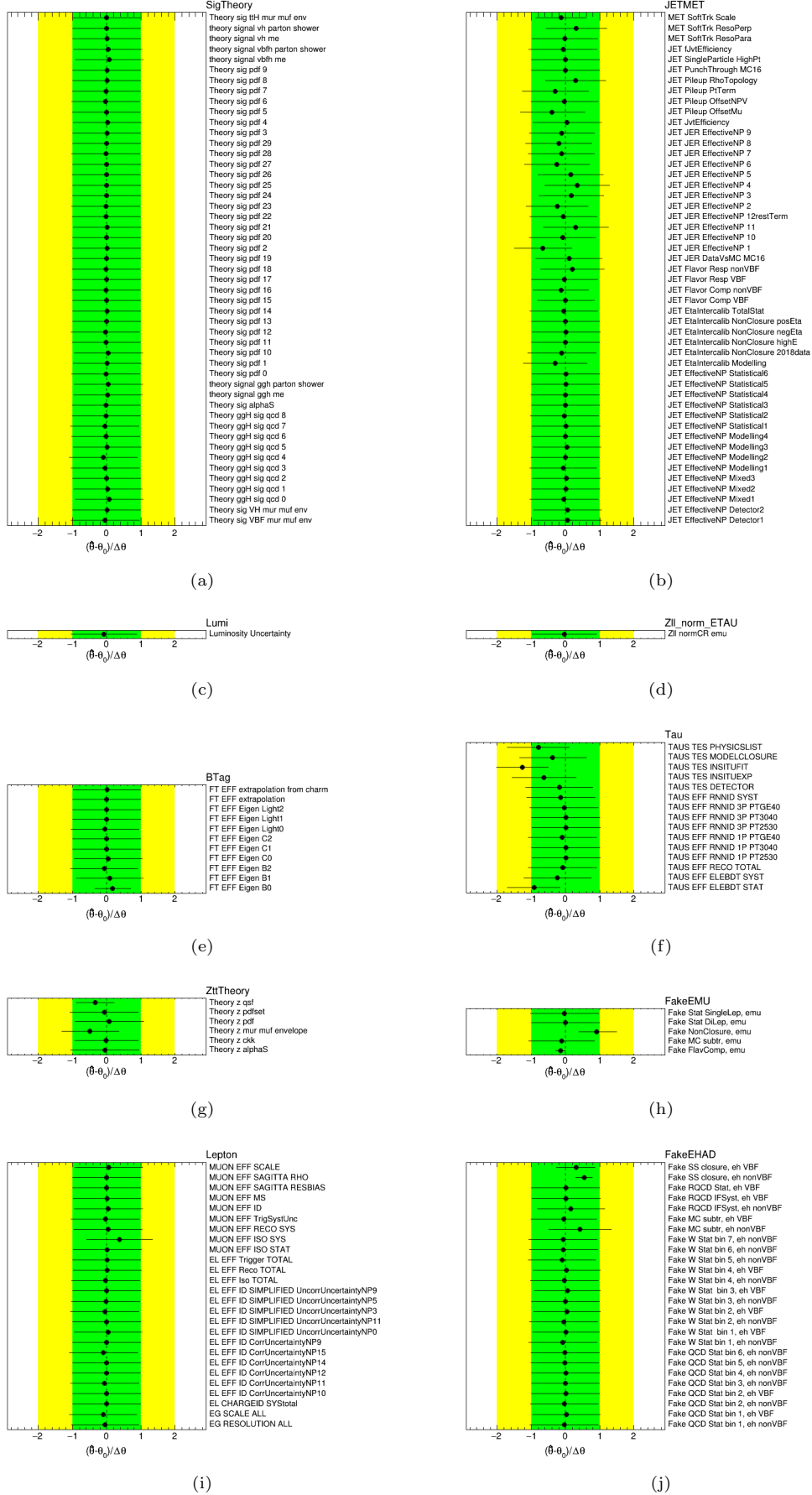


Figure 14.1: The pull distributions of different NPs in the combined $e\tau$ fit (nonVBF+VBF category).

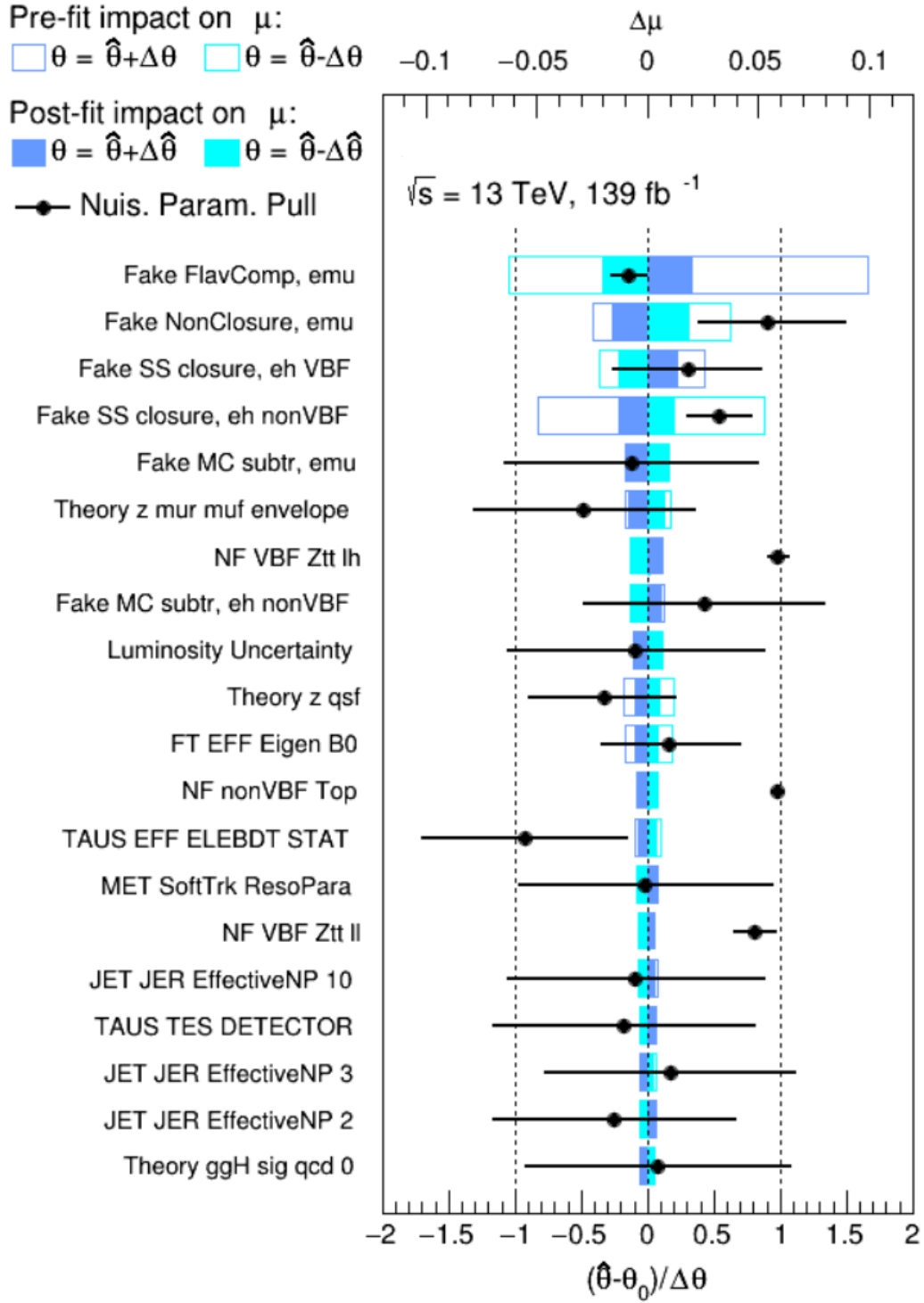


Figure 14.2: The ranking of the top 20 systematic uncertainties with highest post-fit impact on the POI of the combined $e\tau$ fit (nonVBF+VBF).

NF VBF Top	100.0	70.5	75.9	26.9	8.7	20.0	2.7	-2.1	1.4	10.0	6.6	0.9	-26.1	-54.9	8.7	-12.6	0.2	-30.9	-1.5	1.3	0.7	-2.5	-3.4	-0.8	-3.4	-20.3	20.5	8.1
NF VBF Zt lh	70.5	100.0	64.5	3.3	31.6	7.9	13.4	-8.5	1.1	17.9	4.6	-1.2	-14.9	-41.2	8.7	-13.3	4.2	-8.1	7.1	-13.8	-12.3	-2.8	-2.5	-7.0	-1.6	-30.2	8.8	7.7
NF VBF Zt ll	75.9	64.5	100.0	2.0	7.2	9.0	6.9	-25.4	-4.7	27.1	7.0	0.2	-26.2	-38.1	-6.8	-14.1	-4.5	-14.9	3.6	1.5	0.9	-1.8	-2.0	0.0	-2.5	-23.9	5.7	-5.6
NF nonVBF Top	26.9	3.3	2.0	100.0	10.6	87.3	-8.9	39.6	7.7	-20.6	28.7	3.7	3.7	0.9	6.5	4.4	-5.8	-83.8	-1.2	-0.5	-0.2	1.0	-0.0	-0.0	0.5	-5.1	-1.8	21.1
NF nonVBF Zt lh	8.7	31.6	7.2	10.6	100.0	16.0	4.5	-8.9	0.8	14.9	9.2	-0.6	-5.2	-4.6	-6.3	-13.9	-1.4	-9.3	3.7	-54.2	-38.5	-4.2	-9.5	-33.1	3.0	2.9	2.4	26.0
NF nonVBF Zt ll	20.0	7.9	9.0	87.3	16.0	100.0	-4.4	29.0	3.6	-10.9	38.0	3.2	0.2	7.1	-1.8	3.7	-6.9	-75.6	8.1	0.1	0.6	0.7	-1.1	-0.1	0.1	-5.9	-35.1	14.6
μ_{e+}	2.7	13.4	6.9	-8.9	4.5	-4.4	100.0	-9.2	-1.6	35.1	-29.3	-21.6	-2.2	-3.7	2.3	-0.8	0.9	-3.3	-8.0	-2.1	-3.8	-0.6	2.9	1.3	0.3	-2.7	-14.2	-10.0
FT EFF Eigen B0	-2.1	-8.5	-25.4	39.6	-8.9	29.0	-9.2	100.0	-33.1	-41.2	15.6	1.6	4.4	-1.8	27.5	30.1	6.5	-15.1	4.5	0.0	0.4	4.7	7.9	2.7	4.8	-1.2	8.4	12.6
FT EFF Eigen B1	1.4	1.1	-4.7	7.7	0.8	3.6	-1.6	-33.1	100.0	-6.5	2.3	0.4	0.3	0.6	0.5	2.1	-0.6	-1.5	0.6	-0.0	0.0	0.6	0.8	0.3	0.5	-1.2	1.0	4.4
Fake FlavComp, emu	10.0	17.9	27.1	-20.6	14.9	-10.9	35.1	-41.2	-6.5	100.0	-16.1	-2.3	-3.1	-0.2	-12.2	-8.3	-13.5	-6.3	-2.9	-0.3	-1.2	-0.2	2.9	1.8	0.6	-7.0	5.0	12.8
Fake NonClosure, emu	6.6	4.6	7.0	28.7	9.2	38.0	-29.3	15.6	2.3	-16.1	100.0	8.1	-0.5	6.5	-1.4	1.0	-4.5	6.8	2.4	0.3	0.8	0.7	-0.4	0.0	0.3	-7.3	2.9	22.7
Fake SS closure, eh nonVBF	0.9	-1.2	0.2	3.7	-0.6	3.2	-21.6	1.6	0.4	-2.3	8.1	100.0	1.8	-3.0	-0.2	5.3	-27.4	-1.6	38.9	0.6	-1.8	-2.1	-2.4	-1.7	-4.4	4.9	2.9	2.9
JET Etalntcalib Modelling	-26.1	-14.9	-26.2	3.7	-5.2	0.2	-2.2	4.4	0.3	-3.1	-0.5	1.8	100.0	-19.5	-7.3	-10.4	0.0	1.7	-0.6	-0.2	0.0	2.1	4.7	1.3	3.0	4.8	-3.0	-2.3
JET Flavor Comp VBF	-54.9	-41.2	-38.1	0.9	-4.6	7.1	-3.7	-1.8	0.6	-0.2	6.5	-3.0	-19.5	100.0	-7.3	-7.6	-4.9	-2.7	4.7	-0.6	-0.3	1.9	3.0	1.4	2.2	8.1	-9.8	1.1
JET Flavor Comp nonVBF	8.7	8.7	-6.8	6.5	-6.3	-1.8	2.3	27.5	0.5	-12.2	-1.4	-0.2	-7.3	11.7	100.0	-26.3	-2.5	5.3	-3.4	0.0	0.2	-0.7	-1.8	-1.4	-0.9	-1.9	7.6	3.4
JET Pileup RhoTopology	-12.6	-13.3	-14.1	4.4	-13.9	3.7	-0.8	30.1	2.1	-8.3	1.0	5.3	-10.4	-7.6	-26.3	100.0	0.8	1.8	4.0	0.1	0.4	7.4	14.7	4.8	8.3	1.3	-0.1	7.0
MET SoftTrk Scale	0.2	4.2	-4.5	-5.8	-1.4	-6.9	0.9	6.5	-0.6	-13.5	-4.5	-27.4	0.0	-4.9	-2.5	0.8	100.0	9.1	2.3	-1.4	-2.2	-2.2	-5.3	0.1	-4.2	-3.2	-2.7	3.4
MUON EFF ISO SYS	-30.9	-8.1	-14.9	-83.8	-9.3	-75.6	-3.3	-15.1	-1.5	-6.3	6.8	-1.6	1.7	-2.7	5.3	1.8	9.1	100.0	3.5	0.6	0.6	0.1	1.0	0.3	0.5	3.6	2.1	-13.8
TAUS EFF ELEBDT STAT	-1.5	7.1	3.6	-1.2	3.7	8.1	-8.0	4.5	0.6	-2.9	2.4	38.9	-0.6	4.7	-3.4	4.0	2.3	3.5	100.0	-3.6	-4.5	0.1	-1.2	-3.7	-0.9	-0.2	-22.0	-6.5
TAUS EFF RECO TOTAL	1.3	-13.8	1.5	-0.5	-54.2	0.1	-2.1	0.0	-0.0	-0.3	0.3	0.6	-0.2	-0.6	0.0	0.1	-1.4	0.6	-3.6	100.0	-0.8	-0.3	-1.3	-0.6	-0.7	0.7	-0.9	0.0
TAUS EFF RNNID SYST	0.7	-12.3	0.9	-0.2	-38.5	0.6	-3.8	0.4	0.0	-1.2	0.8	-1.8	0.0	-0.3	0.2	0.4	-2.2	0.6	-4.5	-0.8	100.0	-0.9	-2.5	-0.8	-1.4	0.8	-1.7	-0.5
TAUS TES INSITUEXP	-2.5	-2.8	-1.8	1.0	-4.2	0.7	-0.6	4.7	0.6	-0.2	0.7	-2.1	2.1	1.9	-0.7	7.4	-2.2	0.1	0.1	-0.3	-0.9	100.0	-32.0	-8.8	-17.4	-0.8	0.7	0.3
TAUS TES INSITUFT	-3.4	-2.5	-2.0	-0.0	-9.5	-1.1	2.9	7.9	0.8	2.9	-0.4	-2.4	4.7	3.0	-1.8	14.7	-5.3	1.0	-1.2	-1.3	-2.5	-32.0	100.0	-19.5	-39.0	0.2	3.6	2.1
TAUS TES MODELCLCLOSURE	-0.8	-7.0	0.0	-0.0	-33.1	-0.1	1.3	2.7	0.3	1.8	0.0	-1.7	1.3	1.4	-1.4	4.8	0.1	0.3	-3.7	-0.6	-0.8	-8.8	-19.5	100.0	-10.4	-0.2	1.2	1.6
TAUS TES PHYSICSLIST	-3.4	-1.6	-2.5	0.5	3.0	0.1	0.3	4.8	0.5	0.6	0.3	-4.4	3.0	2.2	-0.9	8.3	-4.2	0.5	-0.9	-0.7	-1.4	-17.4	-39.0	-10.4	100.0	-0.5	1.2	0.4
Theory z ckk	-20.3	-30.2	-23.9	-5.1	2.9	-5.9	-2.7	-1.2	-1.2	-7.0	-7.3	4.9	4.8	8.1	-1.9	1.3	-3.2	3.6	-0.2	0.7	0.8	-0.8	0.2	-0.2	-0.5	100.0	-9.1	48.4
Theory z mur muf envelope	20.5	8.8	5.7	-1.8	2.4	-35.1	-14.2	8.4	1.0	5.0	2.9	2.9	-3.0	-9.8	7.6	-0.1	-2.7	2.1	-22.0	-0.9	-1.7	0.7	3.6	1.2	1.2	-9.1	100.0	38.6
Theory z qsf	8.1	7.7	-5.6	21.1	26.0	14.6	-10.0	12.6	4.4	12.8	22.7	2.9	-2.3	1.1	3.4	7.0	3.4	-13.8	-6.5	0.0	-0.5	0.3	2.1	1.6	0.4	48.4	38.6	100.0

Figure 14.3: The nuisance parameter correlation matrix for the combined $e\tau$ fit (nonVBF+VBF).

The setup of the 1-POI fit performed for the $H \rightarrow e\tau$ search by combining the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ final states is summarised in table 13.1. The independent search combines the $\ell\tau_{had}$ channel and the nonVBF category of the $\ell\tau_{\ell'}$ channel from the MC-template method and the VBF category of the $\ell\tau_{\ell'}$ channel from the symmetry method. The resulting upper limits at 95% CL obtained from the standalone fits and the combined fits are shown in figure 14.4, the best-fit values of branching ratios $\hat{B}$ are also shown, and the numerical results with breakdown of statistical uncertainties and systematic uncertainties are provided in [173]. In the $H \rightarrow e\tau$ search, no significance signal is observed, and a 2.2σ excess is found, which is driven mainly by the nonVBF category of the $\ell\tau_{\ell'}$ channel.

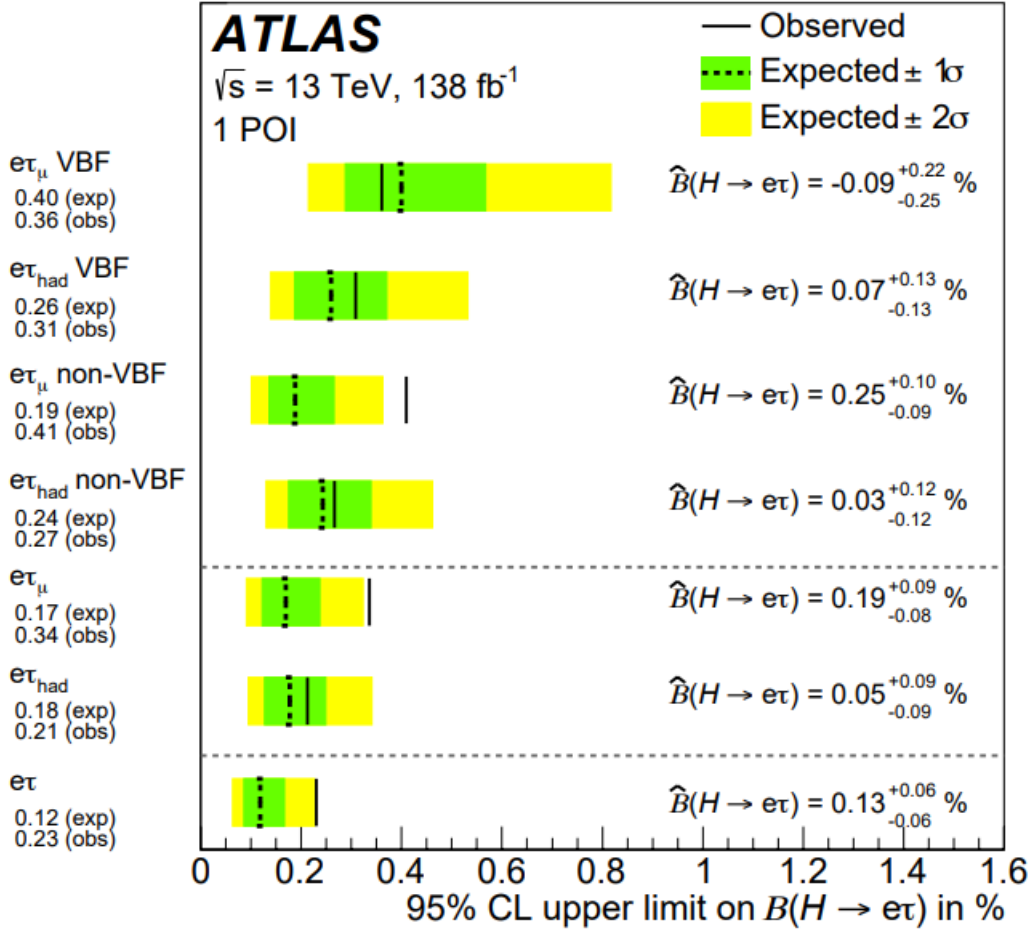


Figure 14.4: The observed (solid line) and expected (dashed line) upper limits at 95% CL on the branching ratio $B(H \rightarrow e\tau)$ obtained from the independent $H \rightarrow e\tau$ search. The best-fit values of the branching ratio $\hat{B}$ are also shown [51].

As discussed in section 3.2.1, the non-diagonal Yukawa coupling matrix elements can be derived using the branching ratios of the LFV Higgs boson decay, as shown in equation 3.11. With the observed upper limits at 95% CL and equation 3.11, one can obtain the limit on the coupling matrix elements as: $\sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 0.0014$, while the individual coupling matrix elements $Y_{\tau e}$ and $Y_{e\tau}$ obtained from the independent $H \rightarrow e\tau$ search are shown in figure 14.5. The naturalness limit $|Y_{ij}Y_{ji}| \lesssim \frac{m_i m_j}{v^2}$ (equation 3.22) obtained from the $\tau \rightarrow \ell\gamma$ searches [43] is also shown.

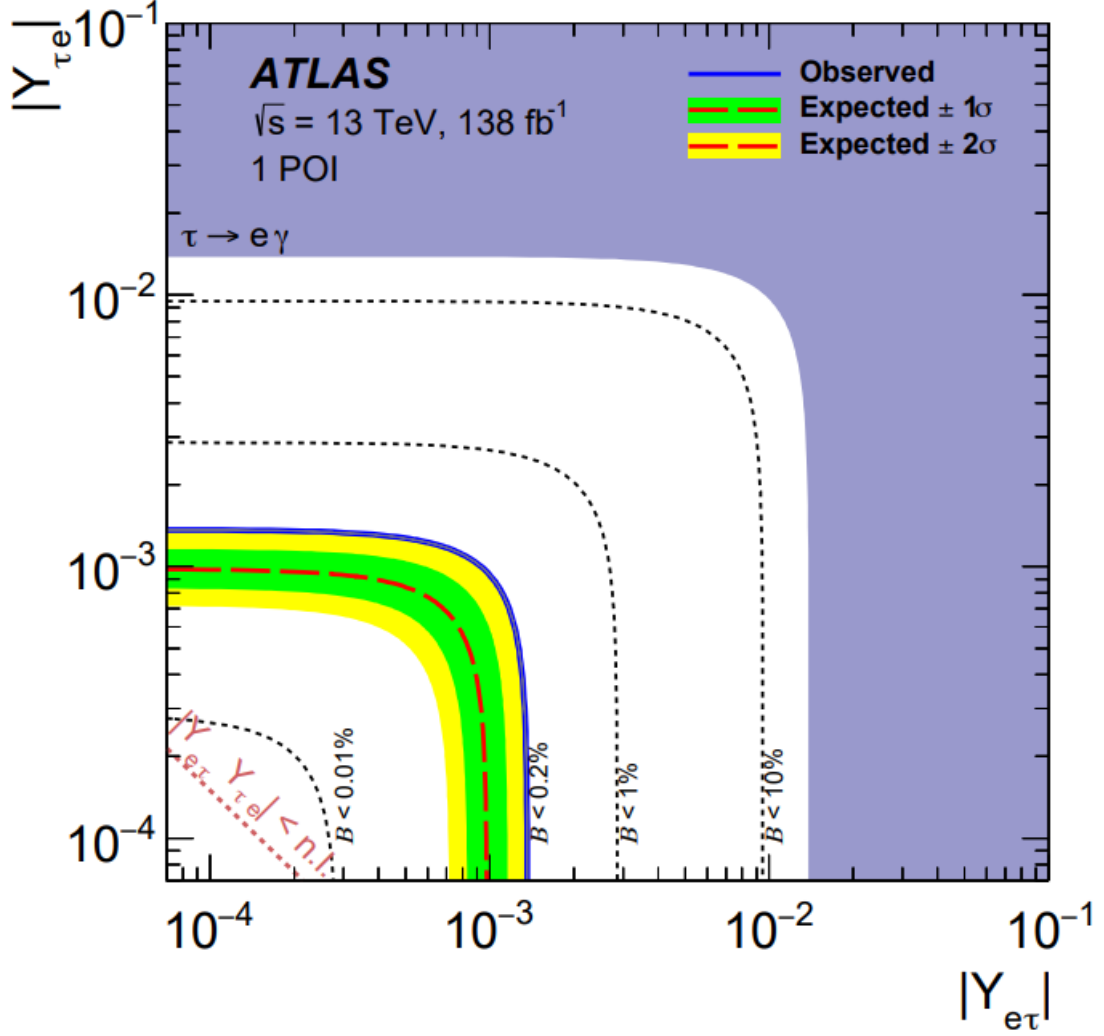


Figure 14.5: The expected (red dashed line) and observed (solid blue line) upper limits at 95% CL from the independent $H \rightarrow e\tau$ search on the absolute values of the individual matrix elements $Y_{\tau e}$ and $Y_{e\tau}$. The constrain from the $\tau \rightarrow \ell\gamma$ searches (dark purple area) and the naturalness limit (dotted red line) are also shown.

14.1.2 I-POI Fits of $H \rightarrow \mu\tau$

In the independent $H \rightarrow \mu\tau$ search, the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ is set to zero. The per-channel and per-category standalone fits of the $H \rightarrow \mu\tau$ search using the MC-template method were also performed. Table 14.3 shows the resulting significance, the upper limits at 95% CL on the $\mathcal{B}(H \rightarrow \mu\tau)$ and the best-fit μ values obtained from each standalone and combined fits.

Similar to the $e\tau_{had}$ standalone fits, the standalone $\mu\tau_{had}$ fits are also performed without the Top normalization factors, but instead including the Top theory uncertainties in the fits as nuisance parameter. The impact of the systematic sources on the POI of the combined $\mu\tau$ fit (nonVBF+VBF) are listed in table 14.4, the ranking of the top 20 uncertainties which have the highest post-fit impact on the POI is shown in figure 14.7. The fake background uncertainties in the $\mu\tau$ channel also have large constraints, which are due to the large value of the prefit uncertainties. The post-fit constraints of the nuisance parameters are shown in figure 14.6. The correlation matrix between the NPs are shown in 14.8.

Results of the Fit with One POI: $H \rightarrow \mu\tau$				
Region	Category	Significance	Upper limit	Best-fit mu value
$\mu\tau_e + \mu\tau_{had}$	nonVBF+VBF	2.5727 (13.11)	0.1941 (0.0895 $^{+0.0359}_{-0.0250}$)	0.1160 $^{+0.0469}_{-0.0456}$
$\mu\tau_e$	nonVBF+VBF	2.0837 (9.45)	0.3078 (0.1630 $^{+0.0644}_{-0.0455}$)	0.1700 $^{+0.0826}_{-0.0817}$
$\mu\tau_e$	nonVBF	0.5160 (7.68)	0.2375 (0.1883 $^{+0.0788}_{-0.0526}$)	0.0517 $^{+0.1021}_{-0.0953}$
$\mu\tau_e$	VBF	0.8507 (3.90)	0.6253 (0.4730 $^{+0.1902}_{-0.1322}$)	0.1989 $^{+0.2390}_{-0.2347}$
$\mu\tau_{had}$	nonVBF+VBF	1.8723 (11.07)	0.1894 (0.1045 $^{+0.0424}_{-0.0292}$)	0.0978 $^{+0.0543}_{-0.0526}$
$\mu\tau_{had}$	nonVBF	1.9182 (9.46)	0.2307 (0.1261 $^{+0.0514}_{-0.0352}$)	0.1203 $^{+0.0654}_{-0.0632}$
$\mu\tau_{had}$	VBF	0.5163 (7.29)	0.2325 (0.1957 $^{+0.0827}_{-0.0547}$)	0.0485 $^{+0.0979}_{-0.0922}$

Table 14.3: The standalone and combined 1-POI fit results of the $H \rightarrow \mu\tau$ search using the MC-template method, the observed (expected) significance, observed (expected) upper limits at 95% CL and the best-fit mu values are shown.

Source of Uncertainty	Impact on POI
Full Unc	+0.047 -0.046
Full Syst	+0.038 -0.037
MC stat	+0.023 -0.023
B-jet	+0.005 -0.005
Fake estimation $\mu\tau_{had}$	+0.014 -0.014
Fake estimation $\mu\tau_e$	+0.012 -0.011
Lepton	+0.005 -0.005
Luminosity	+0.006 -0.005
Jet + MET	+0.010 -0.009
Tau	+0.010 -0.008
Signal Theory Unc.	+0.009 -0.006
Zll normalization	+0.005 -0.006
Z Theory unc.	+0.011 -0.011

Table 14.4: Grouped impact of different systematic sources on the POI of the combined $\mu\tau$ fit (nonVBF+VBF).

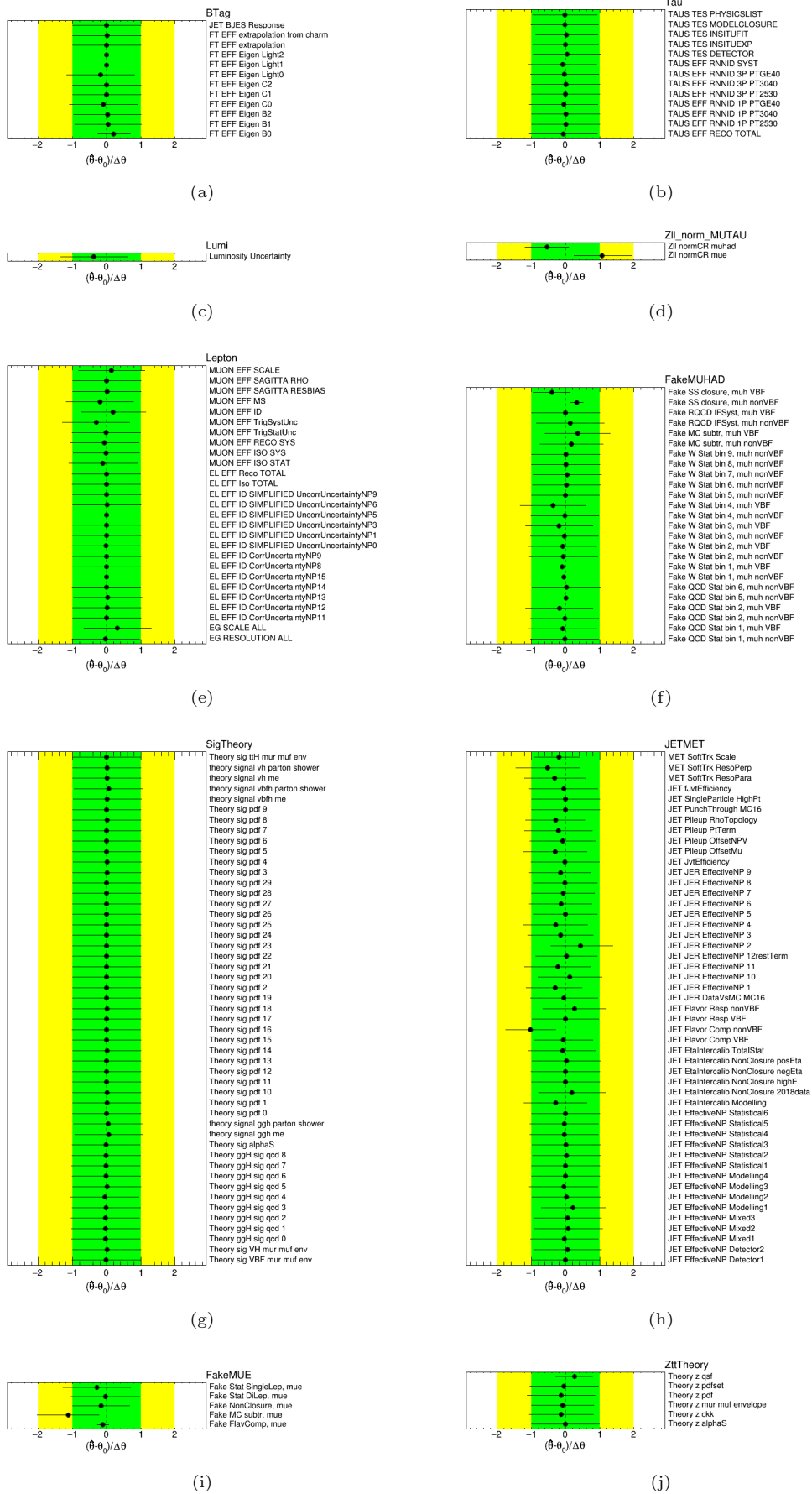


Figure 14.6: The pull distributions of different NPs in the combined $\mu\tau$ fit (nonVBF+VBF category).

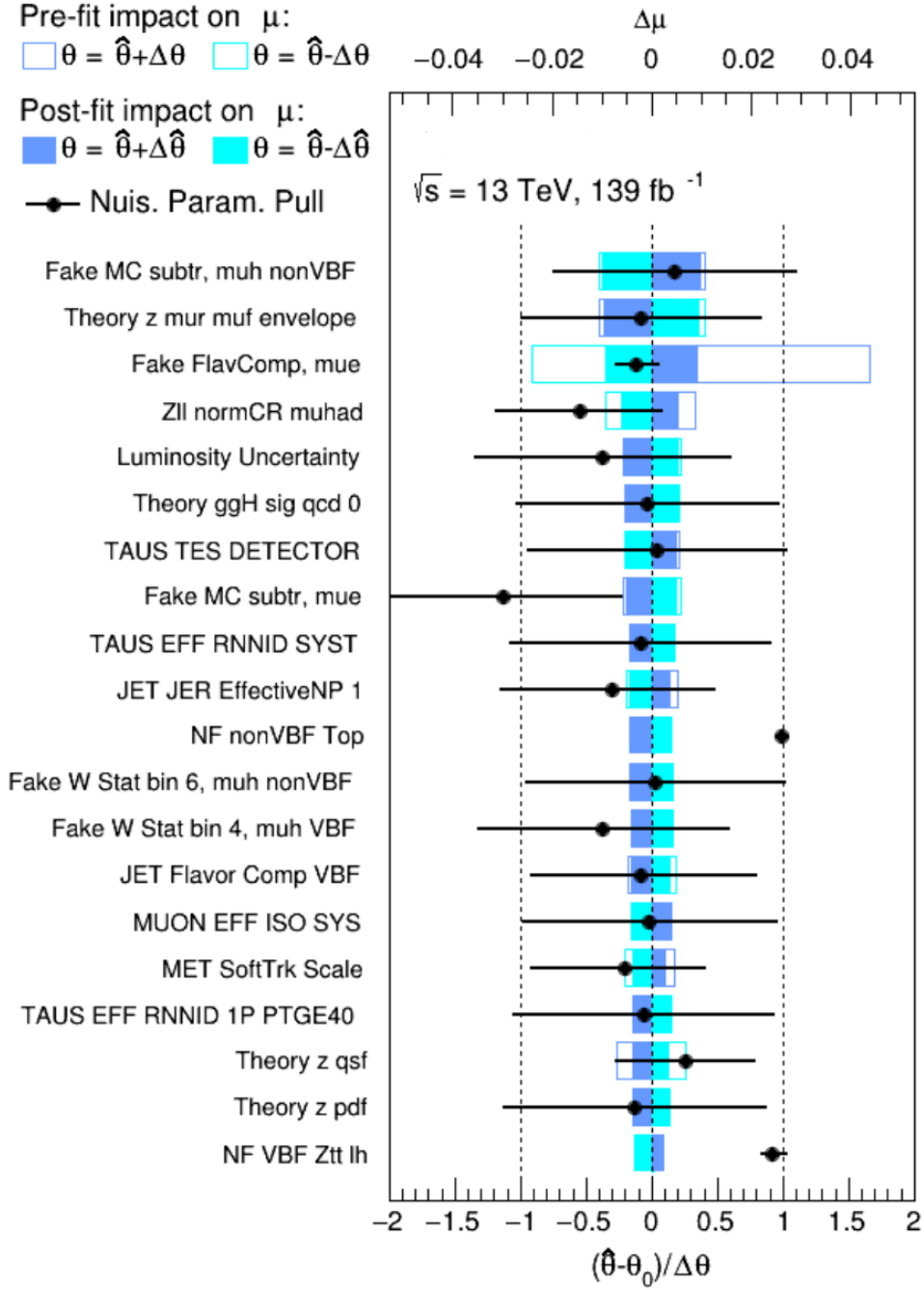


Figure 14.7: The ranking of the top 20 systematic uncertainties with highest post-fit impact on the POI of the combined $\mu\tau$ fit (nonVBF+VBF).

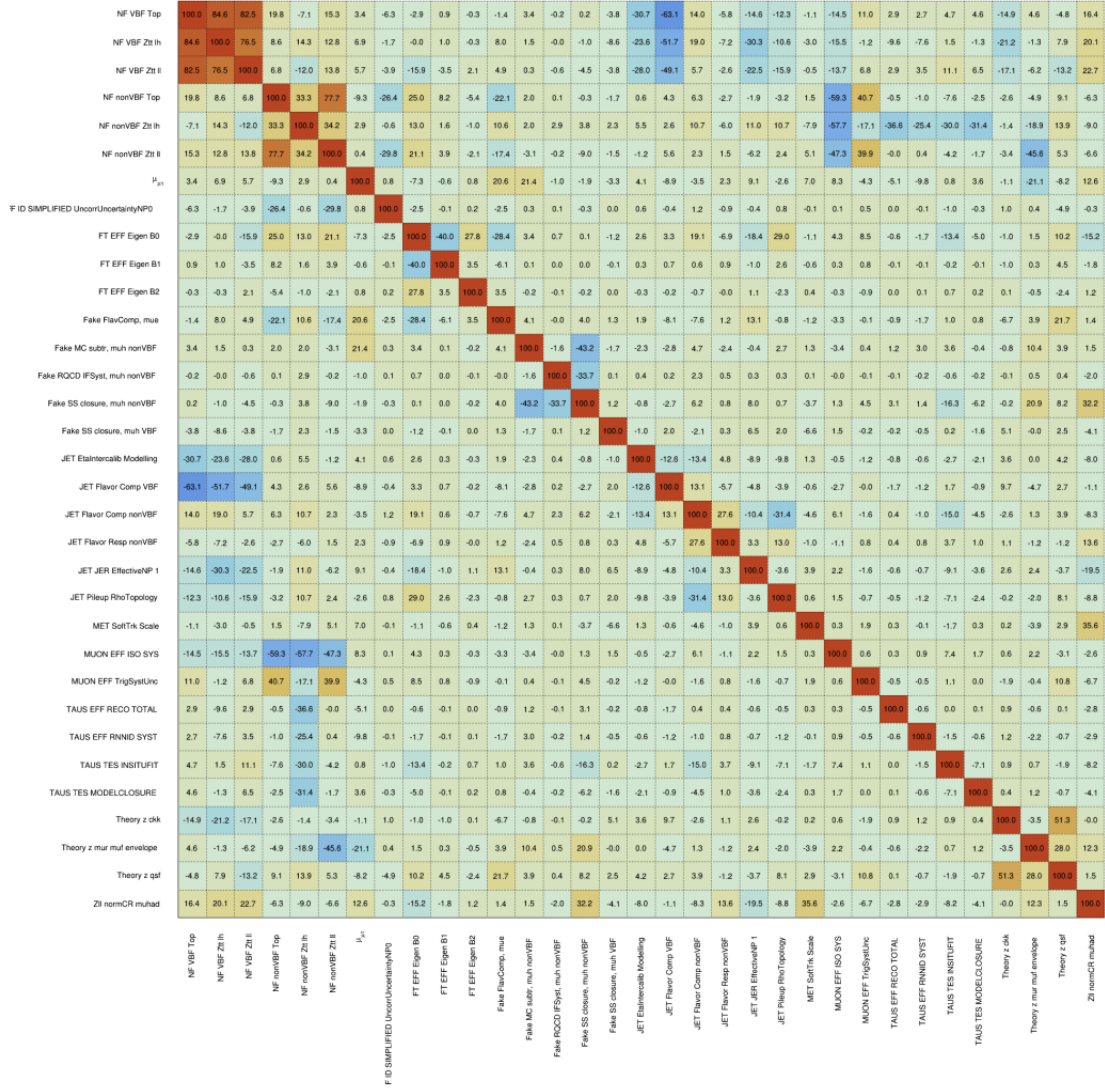


Figure 14.8: The nuisance parameter correlation matrix for the combined $\mu\tau$ fit (nonVBF+VBF).

The combinations of the 1-POI fits performed for the $H \rightarrow \mu\tau$ search, combining the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ final states, are summarized in Table 13.1. Figure 14.9 illustrates the resulting upper limits at 95% confidence level from both standalone and combined fits, along with the corresponding best-fit values of the branching ratio $\hat{\mathcal{B}}$. In this search, an excess of the signal is observed at a significance of 1.9σ , primarily driven by the non-VBF category of the $\ell\tau_{had}$ final state.

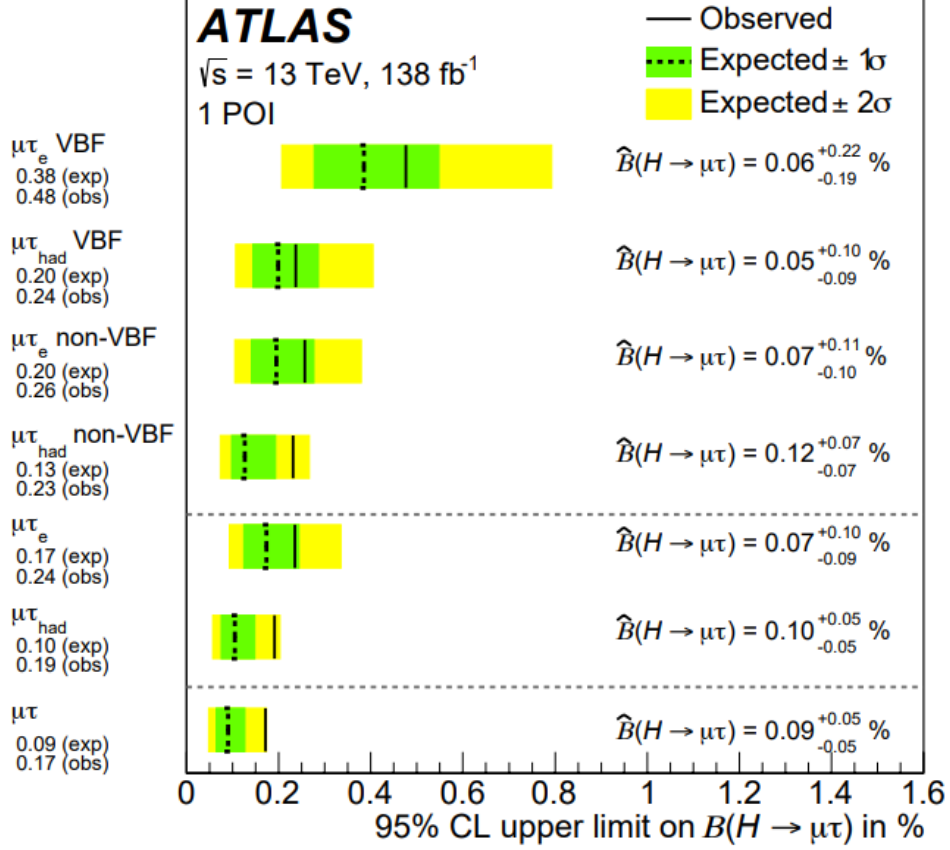


Figure 14.9: The observed (solid line) and expected (dash line) upper limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the independent $H \rightarrow \mu\tau$ search. The best-fit values of the branching ratio $\hat{\mathcal{B}}$ are also shown [51].

For the results of the $H \rightarrow \mu\tau$ search, the corresponding non-diagonal coupling matrix elements are: $\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 0.0012$. Figure 14.10 shows the individual coupling matrix elements $Y_{\mu\tau}$ and $Y_{\tau\mu}$ obtained from the independent $H \rightarrow \mu\tau$ search, the constraints are found to be tighter than the naturalness limit.

14.2 Simultaneous Measurement of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ Signal

The simultaneous 2-POI fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ was performed using the MC-template method, the assumption that the branching ratio of one search is set to zero when fitting the other can be removed. The POIs of the simultaneous fit are the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$, the $e\tau_{\mu}$, $e\tau_{had}$, $\mu\tau_e$ and $\mu\tau_{had}$ final states are fitted simultaneously. This section summarised the 2-POI fit results.

The significance and the best-fit mu values obtained from the simultaneous 2-POI fit are

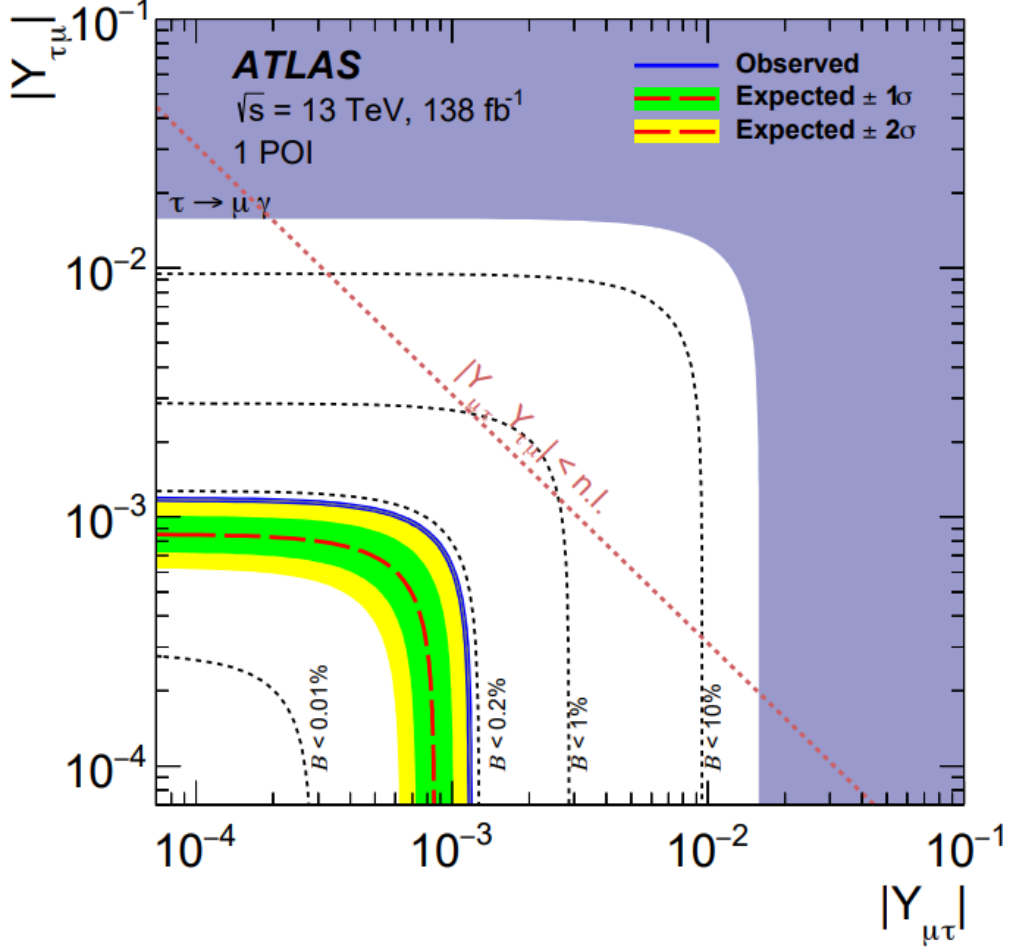


Figure 14.10: The expected (red dash line) and observed (solid blue line) upper limits at 95% CL from the independent $H \rightarrow \mu\tau$ search on the absolute values of the individual matrix elements $Y_{\tau\mu}$ and $Y_{\mu\tau}$. The constrain from the $\tau \rightarrow \ell\gamma$ searches (dark purple area) and the naturalness limit (dotted red line) are also shown [51].

summarised in table 14.5, and the upper limits at 95% CL for the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals are shown in figure 14.11. A 1.6σ excess is observed for the $H \rightarrow e\tau$ signal, which is mainly driven by the nonVBF category of the $\ell\tau_{\ell'}$ final state using the MC-template method. While a 2.4σ excess is observed for the $H \rightarrow \mu\tau$ signal, which is mainly driven by the nonVBF category of the $\ell\tau_{had}$ final state. A likelihood scan has been performed for the two signal strengths, and figure 14.12 shows the likelihood contours at 68% and 95% CL. The compatibility with the Standard Model prediction is found to within 2.1σ .

The non-diagonal matrix elements obtained from the observed upper limits at 95% are: $\sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 0.0013$ and $\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 0.0012$. Figure 14.13 shows the individual coupling matrix elements $Y_{\ell\tau}$ and $Y_{\tau\ell}$ obtained from the simultaneous measurement of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$. Similar to the 1-POI figures, the limits from the $\tau \rightarrow \ell\gamma$ searches and naturalness limit are also provided.

The grouped impact of different systematic sources are shown in table 14.6, and the ranking of the top 20 systematic uncertainties which have highest post-fit impact on each POI is shown in figure 14.14. The fake background uncertainties are also found to have the highest impact on the POIs of the simultaneous fit. The post-fit constraints of the nuisance parameters are shown in

figure 14.15 and figure 14.16, the constraints can be examined when one of the POI is treated as normalization factor in the fit.

The MVA outputs are used as the final discriminants in the statistical analysis, the details of the MVA procedure are described in chapter 11. The BDT score distributions obtained using the MC-template method for the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channels after performing the simultaneous 2-POI fit are shown in figure 14.17 and figure 14.18, respectively.

Region	Category	Significance	Best-fit mu value
POI: $H \rightarrow e\tau$			
combined	nonVBF+VBF	1.6285 (12.04)	$0.0939^{+0.0586}_{-0.0578}$
$\ell\tau_{\mu}$	nonVBF+VBF	1.6695 (9.49)	$0.1354^{+0.0814}_{-0.0810}$
$\ell\tau_{\mu}$	nonVBF	2.0542 (8.22)	$0.1917^{+0.0982}_{-0.0944}$
$\ell\tau_{\mu}$	VBF	-1.1822 (4.18)	$-0.2487^{+0.2102}_{-0.2129}$
$\ell\tau_{had}$	nonVBF+VBF	0.4122 (8.54)	$0.0339^{+0.0842}_{-0.0832}$
$\ell\tau_{had}$	nonVBF	0.3651 (6.60)	$0.0433^{+0.1189}_{-0.1196}$
$\ell\tau_{had}$	VBF	0.2774 (6.04)	$0.0340^{+0.1233}_{-0.1203}$
POI: $H \rightarrow \mu\tau$			
combined	nonVBF+VBF	2.4517 (13.9596)	$0.1071^{+0.0453}_{-0.0441}$
$\ell\tau_{\mu}$	nonVBF+VBF	2.0507 (9.96)	$0.1602^{+0.0786}_{-0.0778}$
$\ell\tau_{\mu}$	nonVBF	1.0585 (8.47)	$0.0972^{+0.0925}_{-0.0915}$
$\ell\tau_{\mu}$	VBF	0.8267 (3.88)	$0.1960^{+0.2414}_{-0.2378}$
$\ell\tau_{had}$	nonVBF+VBF	1.7660 (11.46)	$0.0916^{+0.0538}_{-0.0522}$
$\ell\tau_{had}$	nonVBF	1.7710 (9.72)	$0.1109^{+0.0649}_{-0.0649}$
$\ell\tau_{had}$	VBF	0.7064 (7.23)	$0.0669^{+0.0997}_{-0.0943}$

Table 14.5: The 2-POI fit results of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches using the MC-template method, the observed (expected) significance and the best-fit mu values of each POI are shown.

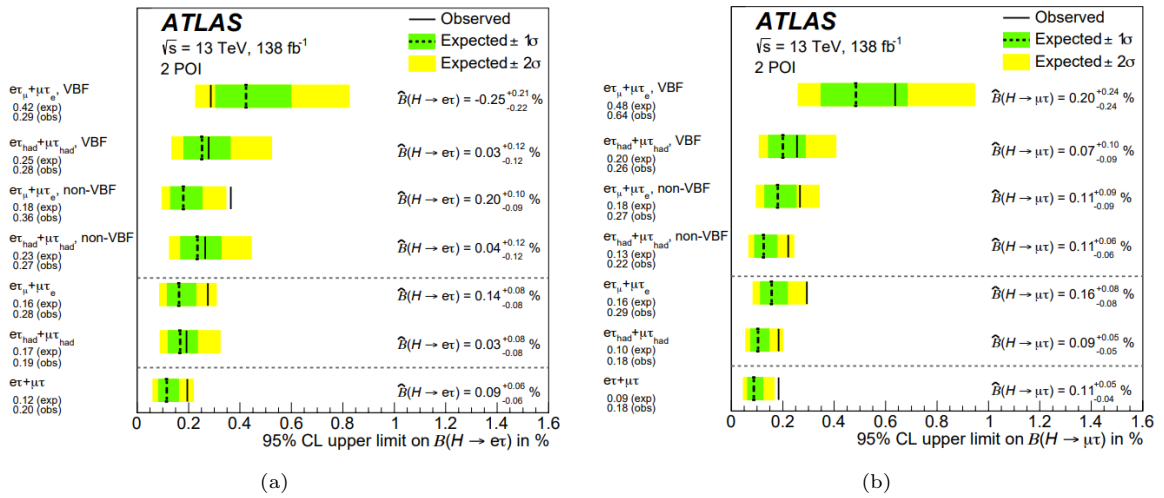


Figure 14.11: The expected (dashed line) and observed (solid line) upper limits at 95% CL on the branching ratio (a) $\mathcal{B}(H \rightarrow e\tau)$ and (b) $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the simultaneous 2-POI fit. The best-fit values of the branching ratios $\hat{\mathcal{B}}$ are also provide [51].

POI	$H \rightarrow e\tau$	$H \rightarrow \mu\tau$
Full Unc	+0.059 -0.058	+0.045 -0.044
Full Syst	+0.051 -0.050	+0.036 -0.035
MC stat	+0.036 -0.037	+0.023 -0.023
B-jet	+0.059 -0.058	+0.045 -0.044
Fake $e\tau_{\text{had}}$	+0.059 -0.058	+0.045 -0.044
Fake $e\tau_{\mu}$	+0.027 -0.027	+0.003 -0.002
Fake $\mu\tau_{\text{had}}$	+0.005 -0.006	+0.015 -0.015
Fake $\mu\tau_e$	+0.009 -0.008	+0.011 -0.010
Lepton	+0.013 -0.013	+0.005 -0.005
Lumi	+0.008 -0.006	+0.005 -0.004
Jet + Met	+0.012 -0.011	+0.009 -0.009
Sig. Theory	+0.008 -0.005	+0.008 -0.005
Tau	+0.009 -0.008	+0.009 -0.008
Zll norm $e\tau$	+0.001 -0.001	+0.001 -0.001
Zll norm $\mu\tau$	+0.002 -0.001	+0.007 -0.008
Z theory	+0.007 -0.007	+0.009 -0.009

Table 14.6: Grouped impact of different systematic sources for the fit result from the simultaneous 2-POI fit of $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals.

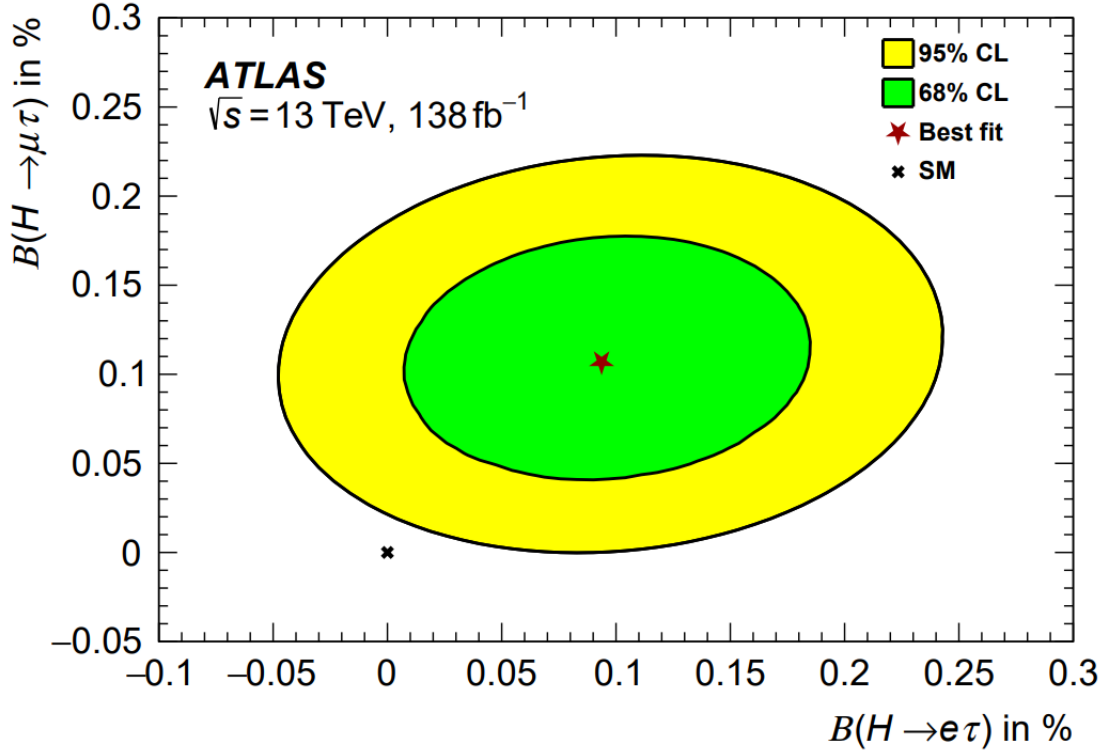


Figure 14.12: The likelihood contours at 68% and 95% CL obtained from the likelihood scan of the simultaneous fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals based on the MC-template method. The best-fit value (read star) of the branching ratios is compared with the SM prediction (black cross) [51].

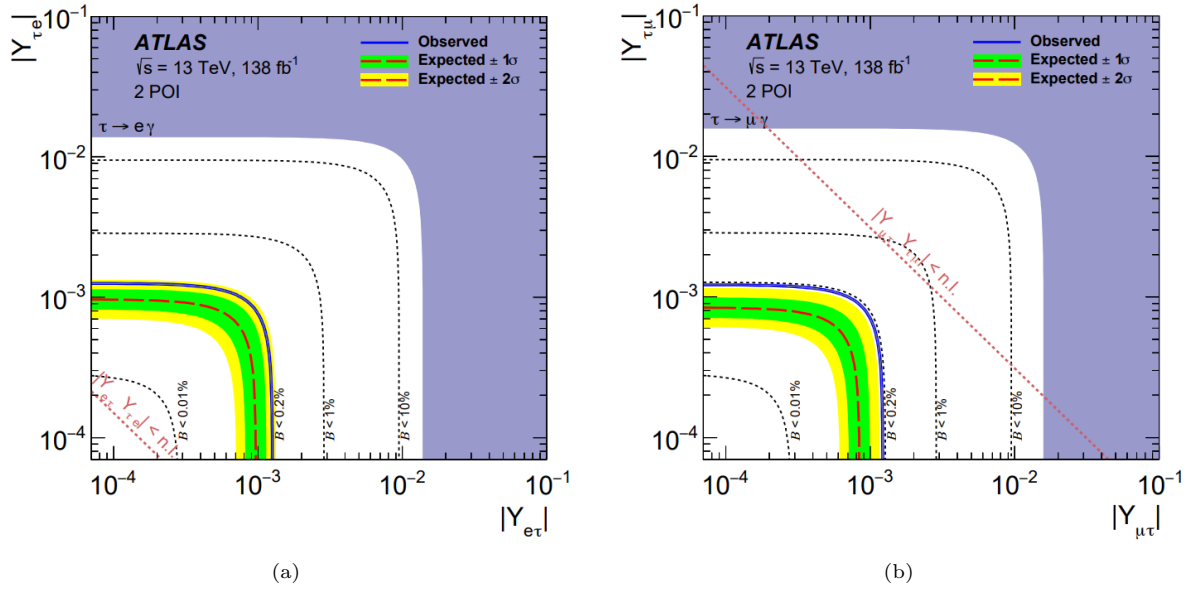


Figure 14.13: The expected (red dash line) and observed (solid blue line) upper limits at 95% CL from the simultaneous measurement of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches using MC-template method on the absolute values of the individual matrix elements $Y_{\tau\mu}$ and $Y_{\mu\tau}$. The constrain from the $\tau \rightarrow \ell\gamma$ searches (dark purple area) and the naturalness limit (dotted red line) are also shown [51].

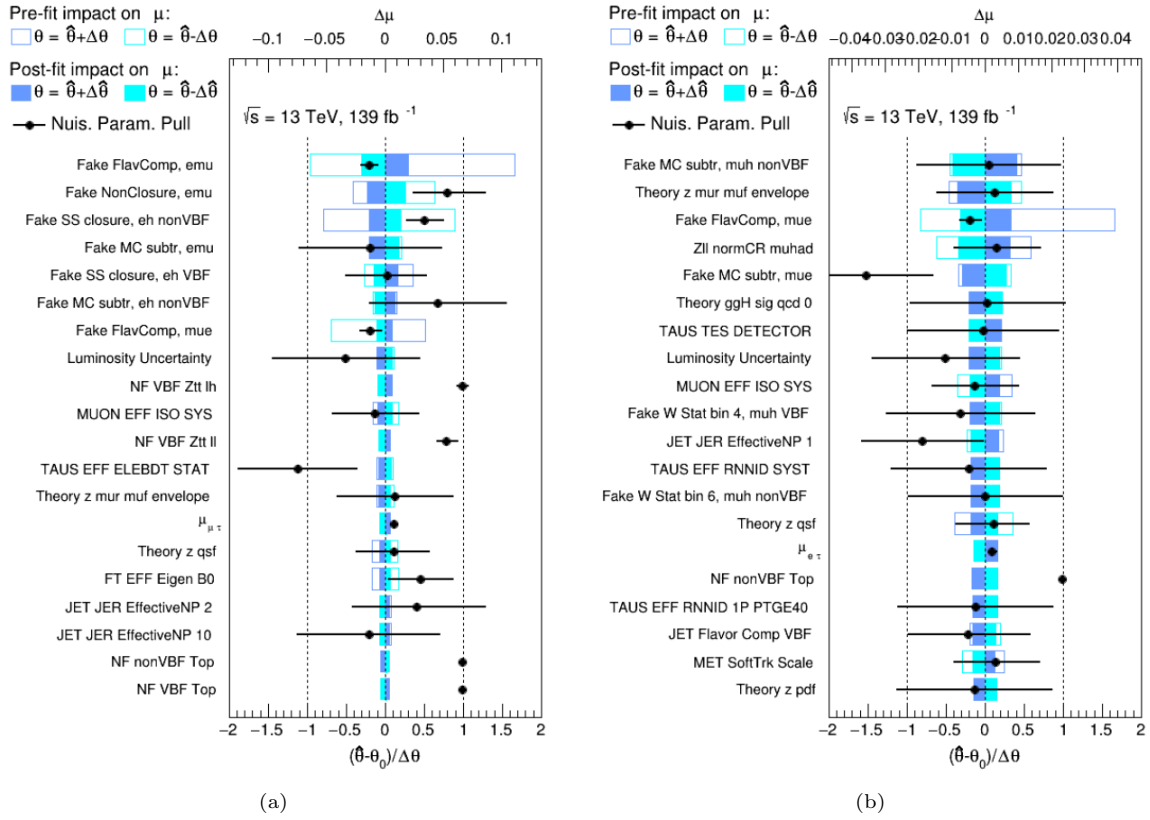


Figure 14.14: The ranking of the top 20 systematic uncertainties with highest post-fit impact on the (a) $\mu_{e\tau}$ and (b) $\mu_{\mu\tau}$ of the simultaneous 2-POI fit.

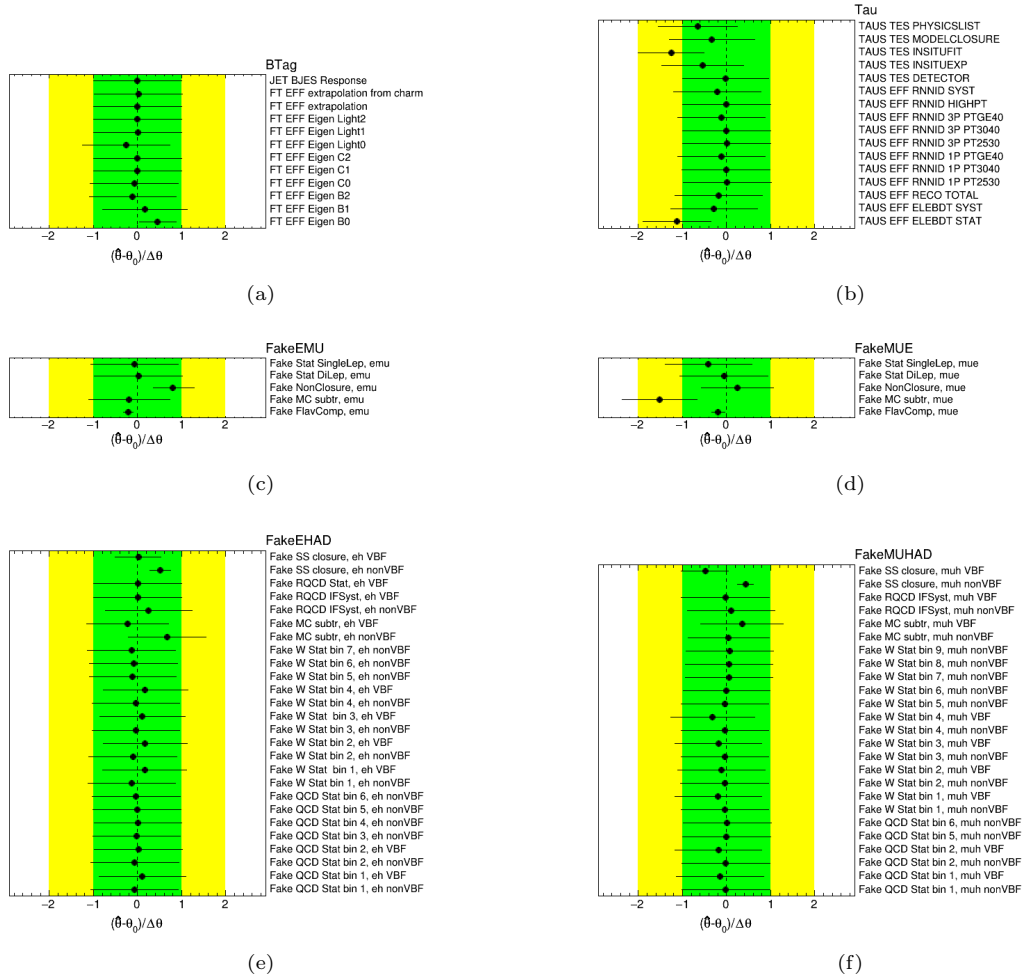


Figure 14.15: The pull distributions of different NPs in the simultaneous 2-POI fit, part1.

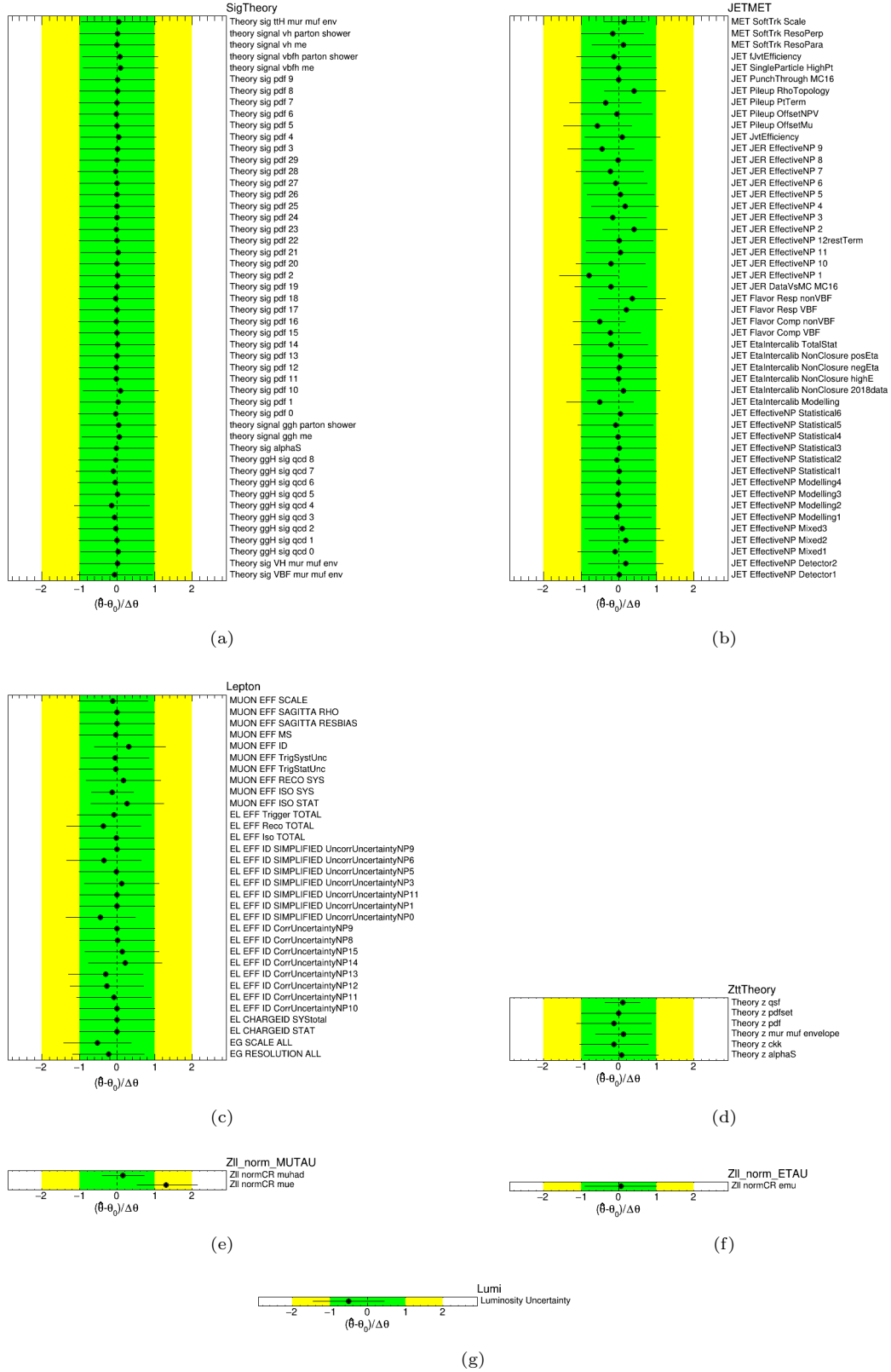


Figure 14.16: The pull distributions of different NPs in the simultaneous 2-POI fit, part2.

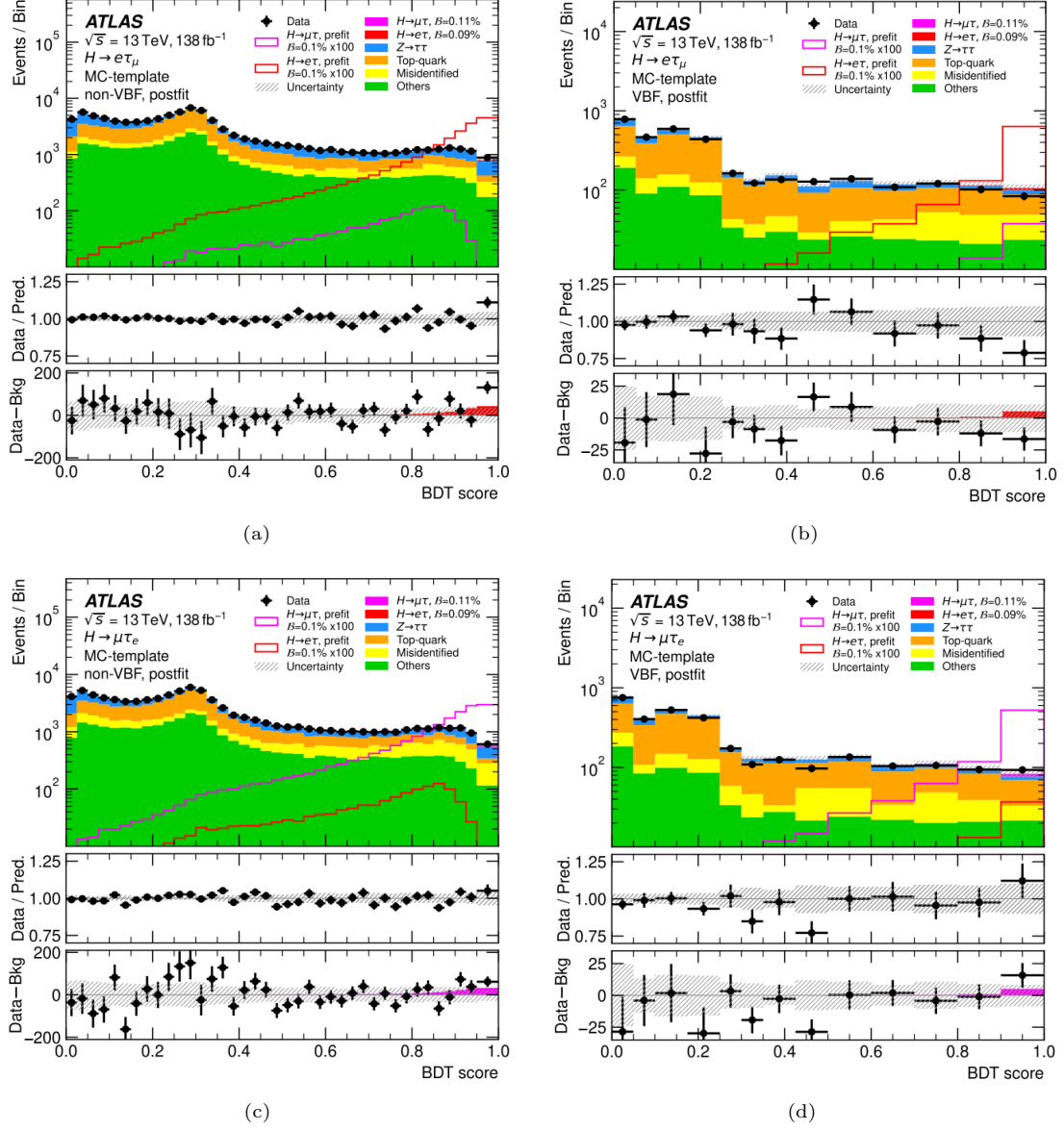


Figure 14.17: The BDT score distributions for the $e\tau\mu$ (top) and $\mu\tau e$ (bottom) final states in the nonVBF (left) and VBF (right) categories of the MC-template method, obtained after performing the simultaneous 2-POI fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals. Overlaid prefit signal shapes asume $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ and are enhanced by a factor of 100 for visibility [51].

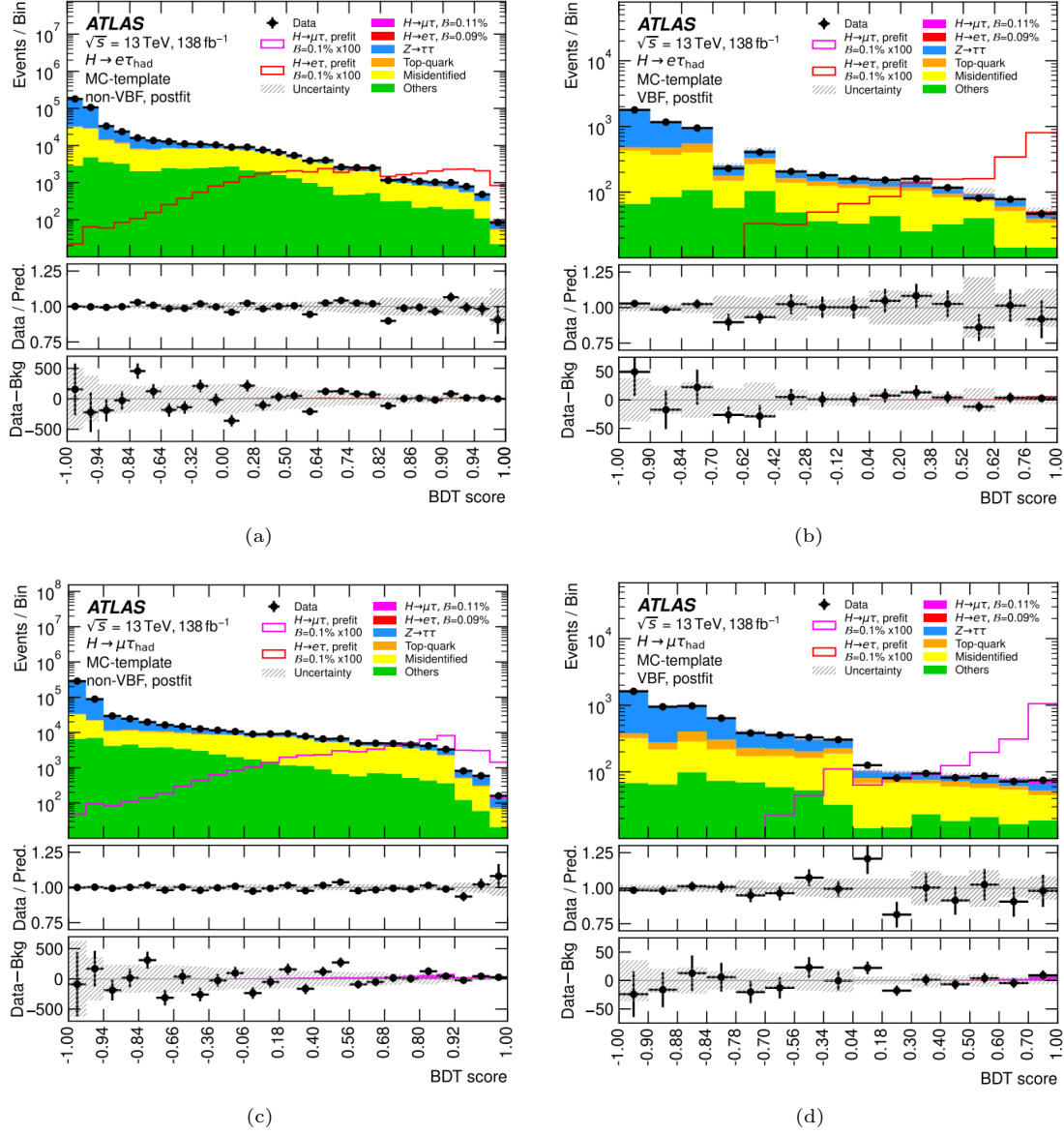


Figure 14.18: The BDT score distributions for the $e\tau_{had}$ (top) and $\mu\tau_{had}$ (bottom) final states in the nonVBF (left) and VBF (right) categories of the MC-template method, obtained after performing the simultaneous 2-POI fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals. Overlaid prefit signal shapes assume $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ and are enhanced by a factor of 100 for visibility [51].

14.3 Measurement of the Branching Ratio Difference

Instead of direct sensitivity to the individual branching ratio of each search, the symmetry method is more sensitive to the branching ratio difference $\mathcal{B}(H \rightarrow e\tau) - \mathcal{B}(H \rightarrow \mu\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$. As shown in equation 10.11, the background in the $\mu\tau_e$ channel is estimated using the $e\tau_\mu$ dataset, and vice versa. Therefore, the results of the symmetry method are presented as the branching ratio difference and the compatibility between the two measurements of the branching ratio difference.

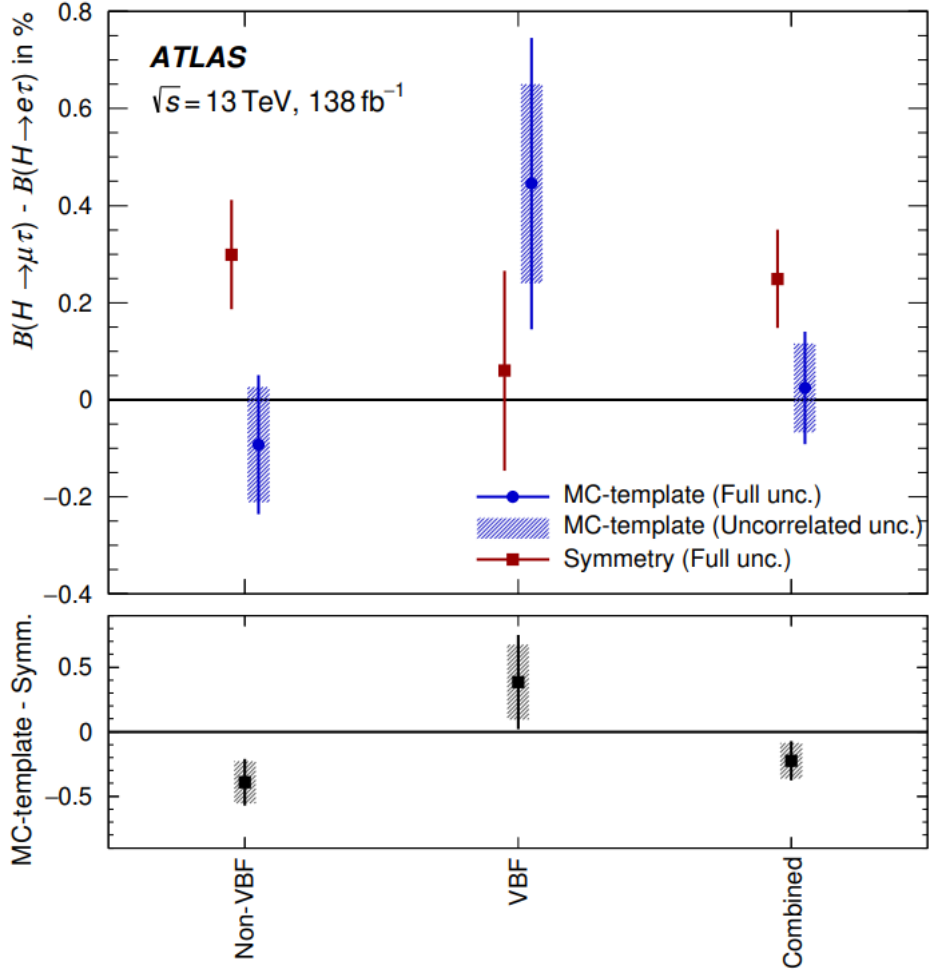


Figure 14.19: The measured best-fit values of $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$, obtained from the standalone fits of the $\ell\tau_{\ell'}$ final state using both symmetry method and MC-template method. In the lower panel, the grey bars represent the uncorrelated uncertainties and the black lines represent the full uncertainties [51].

The branching ratio difference is measured with the results from the standalone 1-POI fits, the simultaneous 2-POI fit in the symmetry method is not considered according to table 13.1. But a 2-POI fit is performed to quantify the compatibility of the best-fit signal strength among the signal regions. The compatibility can be quantified as negative log-likelihood (NLL) values $-2\log \frac{L_{1-POI}}{L_{2-POI}} = 2(NLL_0 - NLL_1)$ [174], in which the 1-POI fit serves as the null-hypothesis, while the 2-POI fit serves as the alternative hypothesis. The NLL values obtained from the 1-POI fit and 2-POI fit are found to be very close to each other in both channel, hence, the best-fit signal strength among the signal regions are compatible with each other.

When performing the 1-POI fits, the branching ratio of one search is assumed to be zero while fitting the other search. The branching ratio difference $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$ is found to have smaller expected uncertainty, therefore, the main result is presented in this form. In the combined fit (nonVBF+VBF), the measured value of the $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$ is $(0.25 \pm 0.10)\%$, which is compatible with zero within 2.5σ .

The branching ratio difference is also measured in each category, the results of the measurements are summarised in figure 14.19. The branching ratio difference for the MC-template method is also measured for checking the compatibility with the symmetry method. The results obtained using MC-template method are also shown in figure 14.19, the $\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$ is measured to be $(0.02 \pm 0.12)\%$ for the combined fit, which is compatible with zero within 1σ . The results obtained with the MC-template and symmetry methods are found to be compatible within 2σ .

Chapter 15

High-Luminosity LHC

In the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches performed using the Run-2 dataset at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 138 fb^{-1} , a small excess around 2σ is observed, as described in chapter 14. The CMS Collaboration also performed the same searches using the MC-template method and reported their results of the observed (expected) upper limits at 95% CL on the branching ratios $\mathcal{B}(H \rightarrow e\tau) < 0.22$ (0.16)% and $\mathcal{B}(H \rightarrow \mu\tau) < 0.15$ (0.15)% [50], which are consistent with the ATLAS results.

The excess of events observed in the Run-2 dataset is not significant enough to claim a discovery, while the Run-2 analysis is limited by the statistical and systematic uncertainties. In order to check the prospect of the future analysis with the upgraded detector technologies, and to estimate the expected results for the searches of a signal among sufficiently larger number of proton-proton collisions, the High-luminosity LHC (HL-LHC) extrapolation procedure is performed by extrapolating the Run-2 dataset.

In this chapter, the study of the HL-LHC scenario is introduced. The procedure for performing the extrapolation of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches from Run-2 to HL-LHC is described in Section 15.1. The extrapolated samples refer to the dataset of proton-proton collisions at $\sqrt{s} = 14$ TeV and correspond to an integrated luminosity of 3000 fb^{-1} . The analysis strategies, signal and background modeling, and selection criteria for the extrapolation study follow the same approach as the Run-2 analysis [51] described in this thesis. Additionally, we discuss the expected reduction in statistical and systematic uncertainties due to the extremely large increase in luminosity. Section 15.2 discusses the expected results obtained from the extrapolation procedure, providing a future insight into the Higgs LFV analysis. The extrapolation results for both the MC-template method and symmetry method are described.

15.1 Extrapolation Procedure

The expected HL-LHC samples are estimated by the extrapolation of the Run-2 analysis samples, the extrapolation procedure is performed by scaling the signal and background samples used in the Run-2 analysis, the scaling factors are defined base on the expected improvements after increasing the luminosity and the centre-of-mass energy.

The large increase in the integrated luminosity from 138 fb^{-1} to 3000 fb^{-1} is accounted by scaling the Run-2 analysis samples by the ratio:

$$SF_{\mathcal{L}} = \frac{\mathcal{L}_{int}(HL - LHC)}{\mathcal{L}_{int}(Run - 2)} = \frac{3000}{138} = 21.74 \quad (15.1)$$

where the integrated luminosity is denoted as $\mathcal{L}_{int}$. While the increase in the centre-of-mass energy $\sqrt{s}$ from 13 TeV to 14 TeV is accounted by scaling the signal and background processes with the scale factors which are defined base on their expected cross-sections¹ and provided in table 15.1, the values of scale factors applied range between 1.10 to 1.21 and they are provided by the ATLAS HiggsProspects [175]. In the symmetry method, the background processes which are symmetric under exchange of the electrons and muons are grouped into "Symm bkg", the corresponding scale factor in table 15.1 takes the mean value of the scale factors corresponding to the individual processes weighted by their respective contributions expected from the MC-simulated samples. The scale factor for the symmetric background is defined as:

$$SF_{\sqrt{s}}^{SymmBkg} = \frac{\sum SF_{process} \times N_{process}}{\sum N_{process}} \quad (15.2)$$

where $N_{process}$ refers to the MC events yield of a process after the baseline selection.

Sample	Scale factor for $\sqrt{s}$	Background estimation method
ggF H	1.12	MC-based+Symmetry-based
VBF H	1.13	MC-based+Symmetry-based
V H	1.10	MC-based+Symmetry-based
ttH	1.21	MC-based
Z+jets	1.10	MC-based
Diboson	1.10	MC-based
Top-quark	1.16	MC-based
W+jets	1.10	MC-based
Fake bkg	1.10	MC-based+Symmetry-based
Symm bkg	1.12	Symmetry-based

Table 15.1: The list of the scale factors applied to different signal and background processes to account for the change in centre-of-mass energy. In each sample, the background estimation method which the corresponding scale factor has been applied is specified.

In the HL-LHC scenario, the detector technologies are expected to be significantly improved in order to meet the challenge of the large increase in the number of interactions per bunch crossing. All the systematic uncertainties described in chapter 12 are expected to be scaled down to account for the potential improvements on the software and hardware in the HL-LHC scenario. The scale factors for each systematic uncertainties are provided by the ATLAS Upgrade Physics group [176], and they are summarised in table 15.2.

¹The scale factors are defined as the ratio of the cross-sections: $SF_{\sqrt{s}} = \frac{\sigma(14 TeV)}{\sigma(13 TeV)}$.

15.1.1 Statistical Uncertainties

The MC statistics in the HL-LHC scenario are expected to be extremely large, rendering the systematic uncertainties arising from limited MC statistics negligible. Therefore, the scale factors for these statistic-related uncertainties are set to zero. As for data-driven backgrounds, the corresponding statistical uncertainties are scaled by a factor to account for the increase in integrated luminosity:

$$SF_{Stat} = \sqrt{\frac{\mathcal{L}_{int}(Run - 2)}{\mathcal{L}_{int}(HL - LHC)}} = \sqrt{\frac{138}{3000}} = 0.21 \quad (15.3)$$

To obtain extrapolation results with more strictly scaled statistical uncertainties, an alternative approach is considered where both MC and data-driven backgrounds are scaled by a factor of 0.21. The results of both approaches are shown in Section 15.2.

15.1.2 Experimental Uncertainties

The experimental uncertainties originate from the detector and the reconstructed objects, the detail description of the experimental uncertainties considered in the Run-2 analysis are given in section 12.2.1.

The large increase in the number of interactions per bunch crossing reduced the statistical components of the uncertainties, but also give rise to a more complex and harsher experimental conditions at the same time. The uncertainties arise from the leptons (electron, muon and $\tau_{had-vis}$) are assumed to have no improvement in the HL-LHC scenario, the scale factor is set to 1. However, for the uncertainties arise from the $\tau_{had-vis}$ identification, the uncertainties arise from the limited size of data are assumed to be negligible and the corresponding scale factor is set to 0.

Most of the uncertainties related to the jets are also scaled by a factor of 1 (no improvement is assumed), except for those uncertainties related to b -/ c -jet tagging and missing transverse energy, the scale factors are set to 0.5 to account for the upgrades in the calorimeters and tracking detectors.

The uncertainty on the luminosity measurement is assumed to be reduced from 1.7% to 1%, the corresponding scale factor is set to 0.59.

15.1.3 Uncertainties in the Misidentified Background

The uncertainties related to the fake background estimation are described in section 12.2.2. The uncertainties related to the statistical components of the fake background estimation are scaled by the factor calculated in equation 15.3 (0.21), while the closure systematic uncertainties are assumed to have no improvement as they are hard to predict.

15.1.4 Uncertainties in the Efficiency Corrections

The uncertainties originate from the efficiency correction procedure for the symmetry method are described in section 12.2.4. The uncertainties related to the statistical components of the lepton efficiency correction are scaled by the factor calculated in equation 15.3 (0.21), while the systematic components are assumed to have no improvement as they are hard to predict.

15.1.5 Theory Uncertainties

The uncertainties originate from the background and signal predictions are described in section 12.2.5. In the MC-templated method, the uncertainties related to the $Z \rightarrow \ell\ell$ and $Z \rightarrow \tau\tau$ background predictions are scaled in two different scheme to account the improvement on the parton density functions of the proton and the Z boson related uncertainties, respectively. The uncertainties related to the PDF of the proton are scaled by 0.4, while the uncertainties related to the Z boson are scaled by 0.5.

As there is no Top control region defined in the $\ell\tau_{had}$ channel, the Top-related theory uncertainties are assigned when performing the standalone fits. The uncertainties related to the PDF and the background modelling are scaled by 0.4, while the others are scaled by 0.5.

As described in section 12.2.5, the theory uncertainties on the signal predictions originate from the SM Higgs decay processes. The scale factors of different production mode of the Higgs boson are summarised in table 15.2, the PDF uncertainties are provided by the conservative approach given in [177] base on the phenomenological predictions.

Uncertainties	Scale factor	Background estimation method
$\tau_{had-vis}$ ID, stat.-related	0.00	MC-based
$\tau_{had-vis}$, others	1.00	MC-based
Electron and muon	1.00	MC-based+Symmetry-based
Flavour tagging c - and b -jets	0.50	MC-based+Symmetry-based
Jet, others	1.00	MC-based+Symmetry-based
E_T^{miss}	0.50	MC-based+Symmetry-based
Fake bkg., stat.-related	0.21	MC-based+Symmetry-based
Fake bkg., others	1.00	MC-based+Symmetry-based
Lepton eff. corr., stat.-related	0.21	Symmetry-based
Lepton eff. corr., others	1.00	Symmetry-based
Z bkg. modelling, PDF	0.40	MC-based
Z bkg. modelling, others	0.50	MC-based
Top-quark bkg. modelling, PDF	0.40	MC-based
Top-quark bkg. modelling, others	0.50	MC-based
Higgs modelling, PDF, ggF	0.41	MC-based+Symmetry-based
Higgs modelling, PDF, VBF H	0.46	MC-based+Symmetry-based
Higgs modelling, PDF, VH	0.46	MC-based+Symmetry-based
Higgs modelling, others	0.50	MC-based+Symmetry-based
Luminosity	0.59	MC-based+Symmetry-based

Table 15.2: The summary of the scale factors applied to the systematic uncertainties described in chapter 12 to extrapolate them from Run-2 to the HL-LHC scenario. In each uncertainties, the background estimation method which the corresponding scale factor has been applied is specified.

15.2 Extrapolation Results

The statistical analysis of the HL-LHC scenario follows the same method described in chapter 13. The statistical fits are performed with the Asimov dataset to obtain the expected sensitivity, the Asimov data is assumed to follow a Poisson distribution. The simultaneous 2-POI fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches is only performed using the MC-template method. For symmetry method, the independent 1-POI fit of one search is performed by assuming the branching ratio of the other search is zero. The results are presented in form of the expected upper limits at 95% CL on the branching ratio by using the CL_s method [170]. Both the results obtained from the nominal approach (statistical uncertainties are scaled by 0) and the alternative approach (statistical uncertainties are scaled by 0.21) are described in this section.

The expected sensitivities obtained from the combined fits of the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channel using the MC-template method are shown in section 15.2.1, both results of the simultaneous 2-POI fit and the individual 1-POI fits are shown. While the expected sensitivities obtained from the independent 1-POI fits of the $\ell\tau_{\ell'}$ channel using the symmetry method are shown in section 15.2.2.

15.2.1 Simultaneous 2-POI Fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ Searches using the MC-template Method

The extrapolation results of the MC-template method are obtained by the simultaneous measurement of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals with the HL-LHC extrapolated samples. Two $Z \rightarrow \tau\tau$ normalization factors and two Top normalization factors are applied to the $Z \rightarrow \tau\tau$ and Top background events, respectively. The normalization factors are included in the fits as free floating parameters.

The expected upper limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the simultaneous 2-POI fit with the HL-LHC extrapolated samples are summarised in table 15.3, both nominal approach and alternative approach results are shown. The comparison on the expected upper limits between the Run-2 analysis results and the HL-LHC results are also shown, in the nominal approach, the expected upper limits at 95% CL are tightened by factors of 0.21 and 0.26 for $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$, respectively, while the upper limits are tightened by factors of 0.24 and 0.27 with the alternative approach.

Approach	$H \rightarrow e\tau$ [%]			$H \rightarrow \mu\tau$ [%]		
	HL-LHC	Run-2	HL-LHC/Run-2	HL-LHC	Run-2	HL-LHC/Run-2
nominal	$0.024^{+0.010}_{-0.007}$	$0.12^{+0.05}_{-0.03}$	0.21	$0.024^{+0.010}_{-0.007}$	$0.09^{+0.04}_{-0.02}$	0.26
alternative	$0.028^{+0.011}_{-0.008}$		0.24	$0.025^{+0.010}_{-0.007}$		0.27

Table 15.3: The summary of the expected upper limits on the $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the simultaneous 2-POI fit using the MC-template method in both the Run-2 and HL-LHC scenario. In the nominal approach, the statistical uncertainties are scaled by 0, while in the alternative approach, the statistical uncertainties are scaled by 0.21 [173, 178].

The expected upper limits at 95% CL on $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from individual channel and categories, as well as their combination are shown in figure 15.1. The comparison between the Run-2 analysis results and the results of the two HL-LHC approaches (nominal and alternative) are also illustrated.

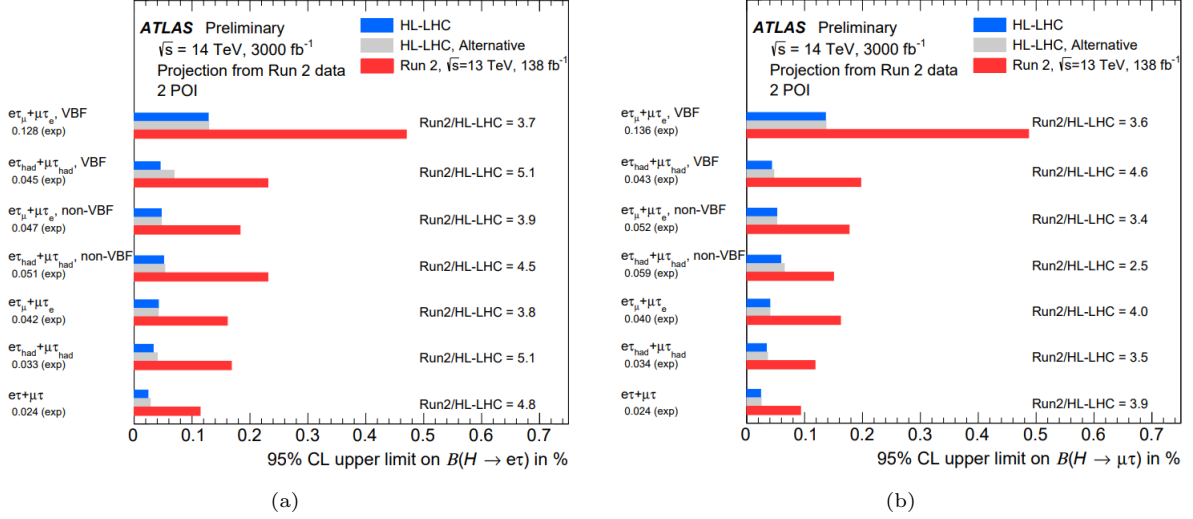


Figure 15.1: The expected upper limits on (a) $\mathcal{B}(H \rightarrow e\tau)$ and (b) $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the simultaneous 2-POI fit of individual channels and categories, as well as their combination. The Run-2 analysis results are included to compare with the results of the two HL-LHC approaches (nominal and alternative) [173, 178].

The expected significance for a signal $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ are shown in figure 15.2, the results of individual channel and categories, as well as their combination are presented. From the combined fits (nonVBF+VBF), the expected significances obtained from the HL-LHC scenario are 8.2σ and 8.1σ for the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signal, respectively, for a signal with branching ratio $\mathcal{B} = 0.1\%$. Comparing with expected significances obtained from the Run-2 analysis, the VBF category of the $\ell\tau_{had}$ channel has the largest improvement with a factor of 5.3. While the improvement for the $\ell\tau_{\ell'}$ channel is between 3.3 and 4.0.

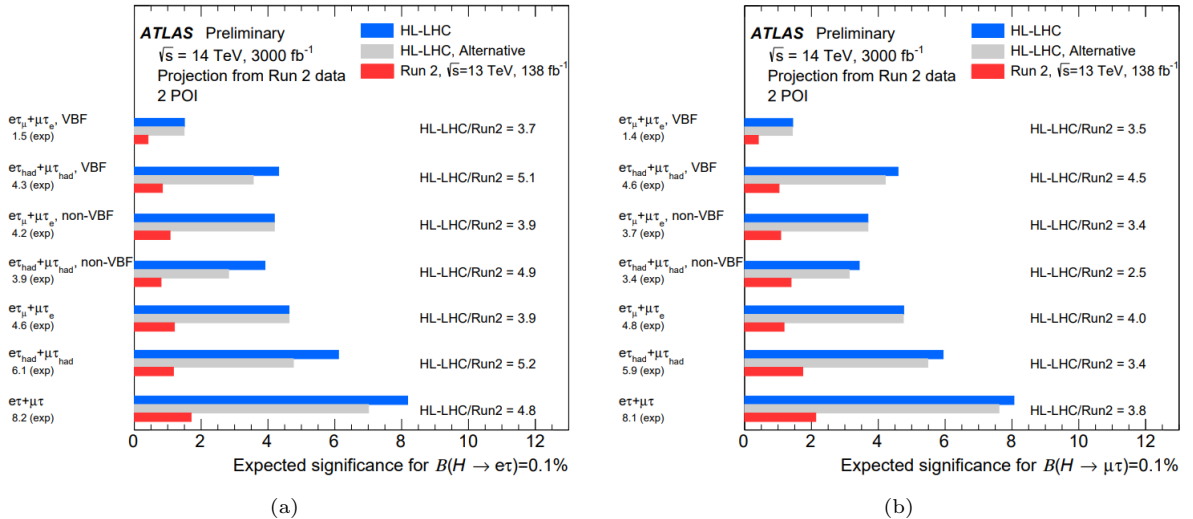


Figure 15.2: The expected significance for a (a) $\mathcal{B}(H \rightarrow e\tau) = 0.1\%$ signal and a (b) $\mathcal{B}(H \rightarrow \mu\tau) = 0.1\%$ signal obtained from the simultaneous 2-POI fit of individual channels and categories, as well as their combination. The Run-2 analysis results are included to compare with the results of the two HL-LHC approaches (nominal and alternative) [173, 178].

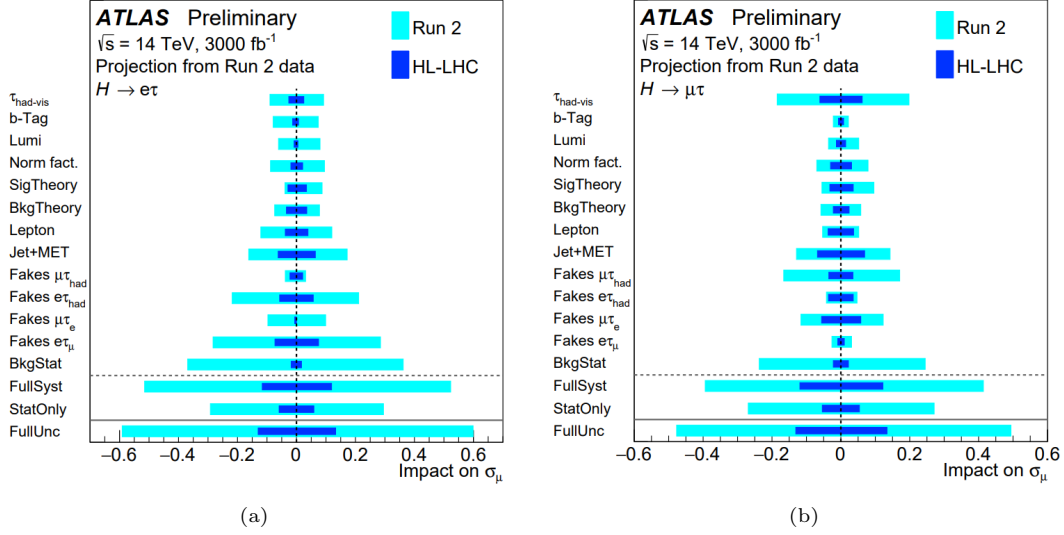


Figure 15.3: Impact of the different types of uncertainties on the signal strength for (a) $\mathcal{B}(H \rightarrow e\tau) = 0.1\%$ and (b) $\mathcal{B}(H \rightarrow \mu\tau) = 0.1\%$, for the combined fit of the MC-template method. Both Run-2 and HL-LHC results are included for comparison.

	$H \rightarrow e\tau$				$H \rightarrow \mu\tau$			
Source of Uncertainty	Run 2		HL-LHC		Run 2		HL-LHC	
Total	+0.599	-0.591	+0.133	-0.130	+0.494	-0.477	+0.134	-0.131
Data statistics	+0.295	-0.292	+0.059	-0.059	+0.271	-0.269	+0.054	-0.054
Full systematic	+0.523	-0.515	+0.119	-0.116	+0.414	-0.394	+0.122	-0.119
Bkg. statistics	+0.362	-0.369	+0.017	-0.018	+0.245	-0.237	+0.022	-0.022
b -tag	+0.074	-0.079	+0.007	-0.013	+0.022	-0.022	+0.008	-0.007
Fake bkg. $e\tau_\mu$	+0.285	-0.283	+0.075	-0.073	+0.031	-0.026	+0.010	-0.009
Fake bkg. $\mu\tau_e$	+0.099	-0.097	< 0.001	-0.006	+0.123	-0.116	+0.058	-0.056
Fake bkg. $e\tau_{had}$	+0.211	-0.218	+0.057	-0.057	+0.047	-0.042	+0.036	-0.036
Fake bkg. $\mu\tau_{had}$	+0.031	-0.038	+0.021	-0.022	+0.171	-0.166	+0.035	-0.035
Jet and E_T^{miss}	+0.172	-0.162	+0.064	-0.062	+0.143	-0.129	+0.069	-0.068
Lepton	+0.120	-0.121	+0.039	-0.038	+0.052	-0.053	+0.037	-0.037
Luminosity	+0.080	-0.061	+0.005	-0.008	+0.052	-0.036	+0.014	-0.013
Norm. factors	+0.095	-0.088	+0.021	-0.019	+0.079	-0.070	+0.031	-0.030
Theory unc. (signal)	+0.087	-0.039	+0.034	-0.029	+0.096	-0.055	+0.036	-0.032
$\tau_{had-vis}$	+0.092	-0.090	+0.025	-0.026	+0.198	-0.185	+0.062	-0.061
Theory unc. (bkg.)	+0.078	-0.074	+0.035	-0.034	+0.058	-0.058	+0.024	-0.022

Table 15.4: The grouped impact of different systematic sources for the simultaneous 2-POI fit results with the Asimov dataset of the $H \rightarrow \ell\tau$ signal using $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$, both the Run-2 and HL-LHC results are included for comparison [178].

Most of the statistical and systematic uncertainties are scaled down in the HL-LHC scenario, their impacts on the signal strength are also reduced compare with the Run-2 analysis results.

Table 15.4 shows the grouped impact of different systematic sources on the signal strength of the simultaneous 2-POI fits for the Run-2 and HL-LHC scenario. The uncertainties on the fake background estimation and the jet and E_T^{miss} are found to have the largest impact on the signal strength in HL-LHC scenario, and the largest reduction are from the statistical uncertainties as expected. The uncertainties originate from the leptons and normalization factors also have large impacts in the Run-2 analysis and largely reduced in the HL-LHC scenario. The direct comparison of the impact of different type of uncertainties between the Run-2 and HL-LHC scenario is shown in figure 15.3. Potential improvements on these uncertainties can be achieved in the future analyses by the use of more refined methods, but the effects are not considered in this study.

15.2.1.1 1-POI fit of the $H \rightarrow e\tau$

The independent 1-POI fit of the $H \rightarrow e\tau$ search is also performed with the Asimov dataset and the HL-LHC extrapolated samples for the MC-template method. The expected significance and expected upper limits at 95% CL on the $\mathcal{B}(H \rightarrow e\tau)$ obtained from standalone per-channel fits are summarised in table 15.5, the Run-2 analysis results are also included for comparison.

The ranking of the top 20 systematic uncertainties with the highest post-fit impact on the signal strength for the HL-LHC scenario is shown in figure 15.4, the ranking plot for the Run-2 analysis is also included for comparing with the HL-LHC results.

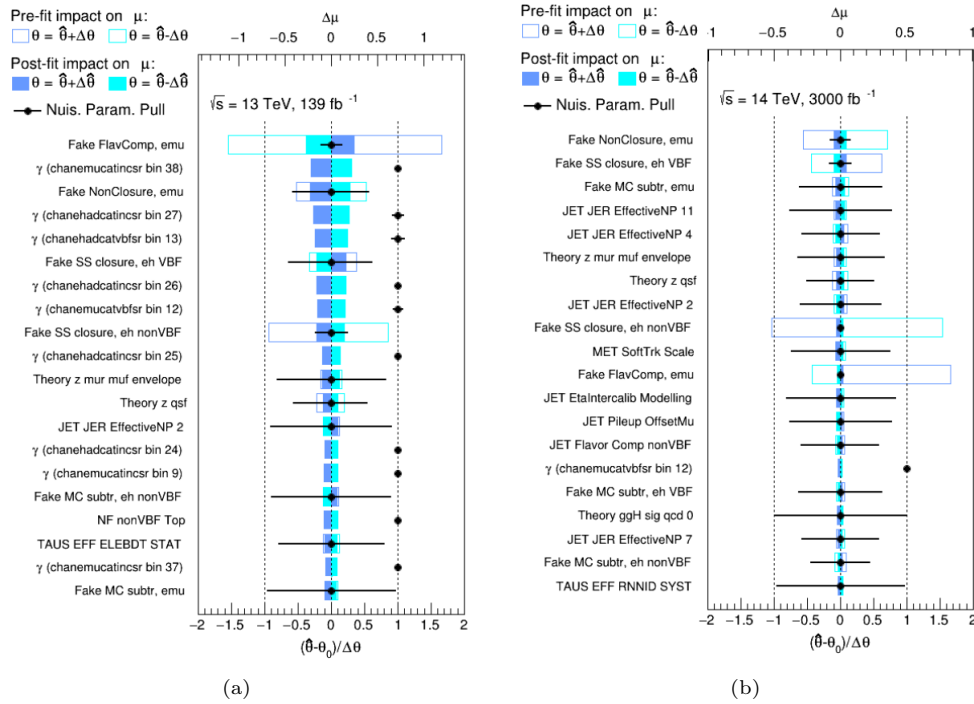


Figure 15.4: The ranking of the top 20 systematic uncertainties with the highest post-fit impact on the POI of the combined fit of $H \rightarrow e\tau$ signal for the (a) Run-2 and (b) HL-LHC scenarios.

15.2.1.2 1-POI fit of the $H \rightarrow \mu\tau$

Similar to the $H \rightarrow e\tau$ search, the independent 1-POI fit of the $H \rightarrow \mu\tau$ search is performed with the Asimov dataset and the HL-LHC extrapolated samples. The expected significance and

Expected 95% CL Upper Limits ($\mathcal{B}(H \rightarrow e\tau)$)			
	$e\tau_\mu$	$e\tau_{had}$	Combined
Run-2	$0.175^{+0.069}_{-0.049}$	$0.175^{+0.071}_{-0.049}$	$0.123^{+0.049}_{-0.035}$
HL-LHC	$0.046^{+0.018}_{-0.013}$	$0.037^{+0.015}_{-0.010}$	$0.028^{+0.011}_{-0.008}$
Expected Significance ($\mathcal{B}(H \rightarrow e\tau) = 0.1\%$)			
Run-2	1.10	1.11	1.57
HL-LHC	4.23	5.31	7.08

Table 15.5: The expected upper limits at 95% CL on $\mathcal{B}(H \rightarrow e\tau)$ and the expected significance for a signal with $\mathcal{B}(H \rightarrow e\tau) = 0.1\%$ obtained from the standalone per-channel fits.

expected upper limits at 95% CL on the $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the standalone per-channel fits are summarised in table 15.6, the Run-2 analysis results are also compared with the HL-LHC extrapolation results.

Figure 15.5 shows the ranking plots of the top 20 systematic uncertainties with the highest post-fit impact on the POI for the Run-2 and HL-LHC scenarios. In both the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches, the uncertainties on the fake background estimation are still expected to have the largest impact, while the jet and missing transverse energy uncertainties are also expected have a large impact in the HL-LHC scenario.

Expected 95% CL Upper Limits ($\mathcal{B}(H \rightarrow \mu\tau)$)			
	$\mu\tau_e$	$\mu\tau_{had}$	Combined
Run-2	$0.167^{+0.064}_{-0.046}$	$0.122^{+0.052}_{-0.034}$	$0.096^{+0.040}_{-0.027}$
HL-LHC	$0.044^{+0.017}_{-0.012}$	$0.037^{+0.015}_{-0.010}$	$0.027^{+0.011}_{-0.008}$
Expected Significance ($\mathcal{B}(H \rightarrow \mu\tau) = 0.1\%$)			
Run-2	1.16	1.70	2.08
HL-LHC	4.30	5.51	7.26

Table 15.6: The expected upper limits at 95% CL on $\mathcal{B}(H \rightarrow \mu\tau)$ and the expected significance for a signal with $\mathcal{B}(H \rightarrow \mu\tau) = 0.1\%$ obtained from the standalone per-channel fits.

15.2.2 Independent 1-POI Fits of the $\ell\tau_{\ell'}$ Final States using Symmetry Method

The extrapolation results of the symmetry method are obtained from the independent 1-POI fits of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals with the Asimov dataset and using the extrapolated samples. The branching ratio of one search is set to zero when performing the fit of the other search. In the symmetry method, only the $\ell\tau_{\ell'}$ channel is considered in the HL-LHC extrapolation procedure, the statistical analysis follows the method described in chapter 13.

The expected upper limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ obtained from the independent 1-POI fits with the Asimov dataset and the HL-LHC extrapolated

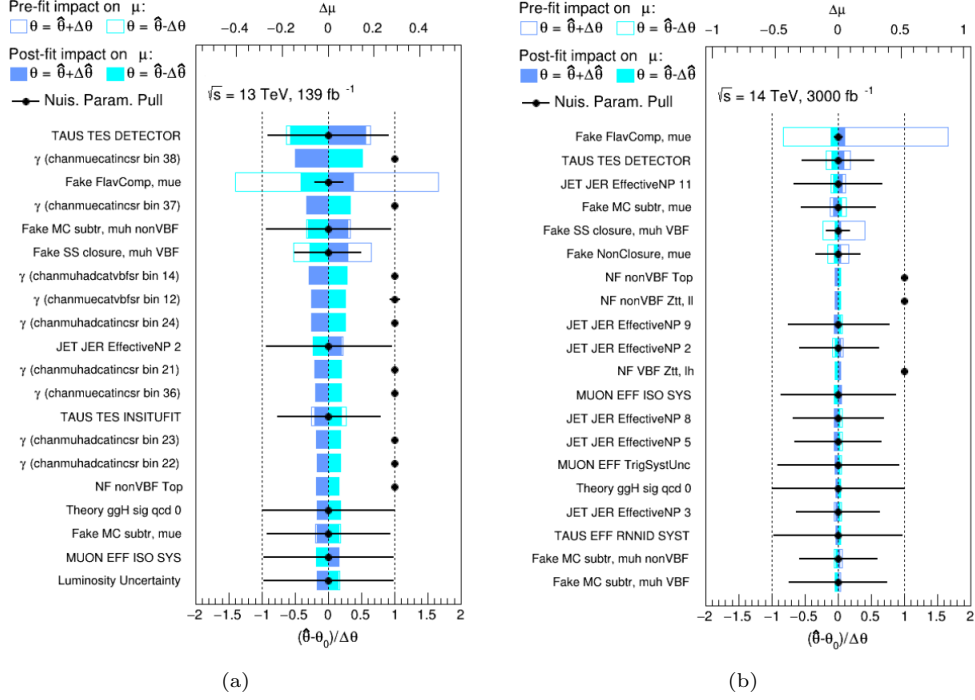


Figure 15.5: The ranking of the top 20 systematic uncertainties with the highest post-fit impact on the POI of the combined fit of $H \rightarrow \mu\tau$ signal for the (a) Run-2 and (b) HL-LHC scenarios.

samples are summarised in table 15.7, the Run-2 analysis results are included to compare with the results of the HL-LHC nominal approach. In the combined fits (nonVBF+VBF), the upper limits at 95% CL are obtained as $0.51^{+0.020}_{-0.014}$ for $H \rightarrow e\tau$ and $0.052^{+0.020}_{-0.014}$ for $H \rightarrow \mu\tau$. A graphical comparison which also include the results of the alternative approach is shown in figure 15.6, the improvements compared to the Run-2 analysis results are in the range of 3.1-4.9.

Expected 95% CL Upper Limits ($\mathcal{B}(H \rightarrow \ell\tau)$)						
	$H \rightarrow e\tau$ [%]			$H \rightarrow \mu\tau$ [%]		
	HL-LHC	Run-2	Run-2/HL-LHC	HL-LHC	Run-2	Run-2/HL-LHC
VBF	$0.070^{+0.028}_{-0.020}$	$0.338^{+0.139}_{-0.095}$	4.83	$0.084^{+0.034}_{-0.024}$	$0.404^{+0.164}_{-0.113}$	4.79
nonVBF	$0.073^{+0.028}_{-0.020}$	$0.230^{+0.092}_{-0.064}$	3.16	$0.067^{+0.027}_{-0.019}$	$0.214^{+0.084}_{-0.060}$	3.22
VBF+nonVBF	$0.051^{+0.020}_{-0.014}$	$0.190^{+0.075}_{-0.053}$	3.70	$0.052^{+0.020}_{-0.014}$	$0.188^{+0.074}_{-0.053}$	3.65

Table 15.7: The expected upper limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ for the Run-2 and HL-LHC scenario. The values of the upper limits obtained from the fits of each individual category and combined fit are included, and the HL-LHC results are compared to the Run-2 analysis results.

The expected significance for a signal with $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ obtained from the independent 1-POI fits of the each category and the combined fit are summarised in table 15.8, both results of the Run-2 and HL-LHC scenario are included and compared. In the Asimov fits performed with HL-LHC extrapolated samples, the expected significance for the $H \rightarrow e\tau$ signal is obtained as 3.6σ , while the expected significance for the $H \rightarrow \mu\tau$ signal is obtained as 3.7σ . Figure 15.7 shows the comparison of the expected significances between the Run-2 and the two HL-LHC approaches

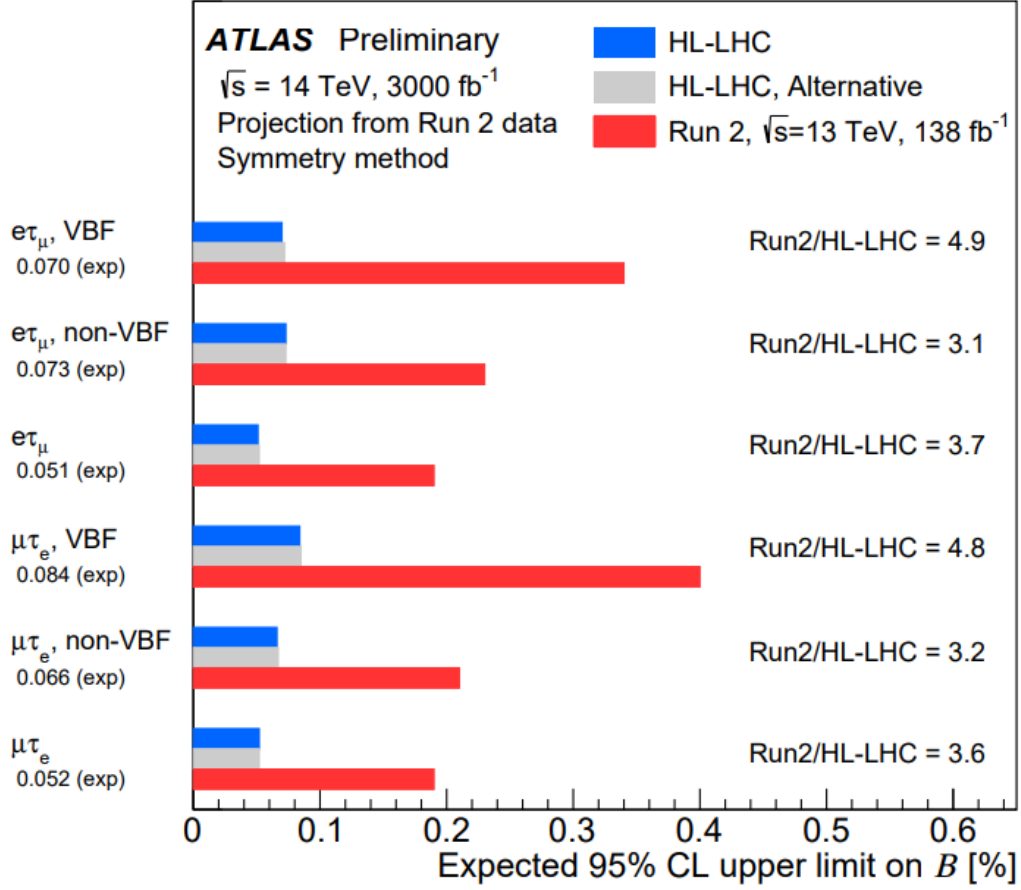


Figure 15.6: The expected upper limits on $\mathcal{B}(H \rightarrow \ell\tau)$ obtained from the independent 1-POI fits with the HL-LHC extrapolated samples, for both the nominal and alternative approach. The Run-2 analysis results are included to compare with the HL-LHC scenario [178].

(nominal and alternative), the largest improvements are obtained in the VBF category as expected. The statistics is low in this region, resulting in higher statistical uncertainties. The improvements on the expected significances in the HL-LHC scenario are within the range of 3.1-4.7.

Expected Significance ($\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$)						
	$H \rightarrow e\tau$ [%]			$H \rightarrow \mu\tau$ [%]		
	HL-LHC	Run-2	Run-2/HL-LHC	HL-LHC	Run-2	Run-2/HL-LHC
VBF	2.57	0.55	4.67	2.23	0.48	4.65
nonVBF	2.60	0.85	3.06	2.96	0.90	3.29
VBF+nonVBF	3.59	1.00	3.67	3.74	1.02	3.67

Table 15.8: The expected significances for a signal with $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ from the symmetry method. The values of the expected significances obtained from the fits of each individual category and combined fit are included, and the HL-LHC results are compared to the Run-2 analysis results.

The impact of different systematic sources on the signal strength in the $e\tau_\mu$ and $\mu\tau_e$ channels for the Run-2 and HL-LHC scenarios is summarized in Table 15.9. The dominant contribution to

the total impact on the signal strength comes from the systematic uncertainties. As expected, the statistical uncertainty is significantly reduced in the HL-LHC scenario. However, in the symmetry method, the statistical uncertainty from the HL-LHC data is still expected to dominate, while the MC-templated method is expected to be more impacted by systematic uncertainties. In the HL-LHC scenario, similar to the MC-template method, the largest impacts originate from uncertainties in the fake background estimation, efficiency correction, jets, and missing transverse energy. Figure 15.8 provides a graphical representation of the impact of different types of uncertainties on the signal strength uncertainties σ_μ for both channels. It should be noted that potential improvements in these uncertainties can be achieved in future analyses through the use of more refined methods, but their effects are not considered in this study.

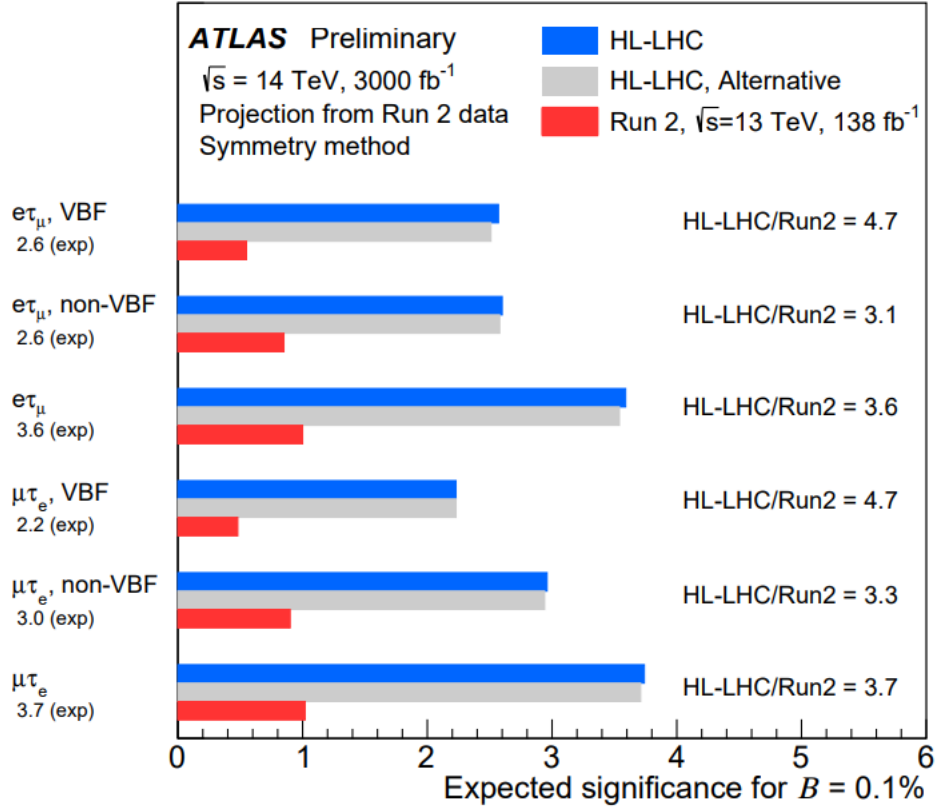


Figure 15.7: The expected significance for a signal with $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$ from the symmetry method, obtained from the independent 1-POI fits of each individual category and the combined fit with the HL-LHC extrapolation samples. The Run-2 analysis results are also shown for comparing with the two HL-LHC approaches (nominal and alternative) [178].

	$H \rightarrow e\tau$				$H \rightarrow \mu\tau$			
Source of Uncertainty	Run 2		HL-LHC		Run 2		HL-LHC	
Total	+0.994	-0.993	+0.280	-0.277	+0.975	-0.981	+0.273	-0.268
Data statistics	+0.508	-0.499	+0.102	-0.101	+0.542	-0.534	+0.109	-0.108
Full systematic	+0.853	-0.859	+0.261	-0.257	+0.814	-0.820	+0.250	-0.246
Bkg. statistics	+0.706	-0.737	+0.177	-0.181	+0.674	-0.695	+0.170	-0.172
b -tag	+0.006	0.000	+0.004	-0.003	+0.004	-0.001	+0.003	-0.002
Fake bkg.	+0.420	-0.451	+0.086	-0.092	+0.479	-0.469	+0.120	-0.105
Jet	+0.140	-0.176	+0.097	-0.105	+0.156	-0.158	+0.115	-0.110
E_T^{miss}	+0.216	-0.145	+0.104	-0.087	+0.059	-0.088	+0.051	-0.041
Lepton	+0.125	-0.142	+0.044	-0.049	+0.158	-0.146	+0.074	-0.066
Luminosity	+0.038	-0.006	+0.015	-0.009	+0.028	-0.003	+0.012	-0.007
Theory unc. (signal)	+0.073	-0.003	+0.024	-0.014	+0.078	-0.003	+0.026	-0.015

Table 15.9: The grouped impact of different systematic sources for the individual 1-POI fit results using symmetry method with the Asimov dataset of the $H \rightarrow \ell\tau$ signal using $\mathcal{B}(H \rightarrow \ell\tau) = 0.1\%$, both the Run-2 and HL-LHC results are included for comparison [178].

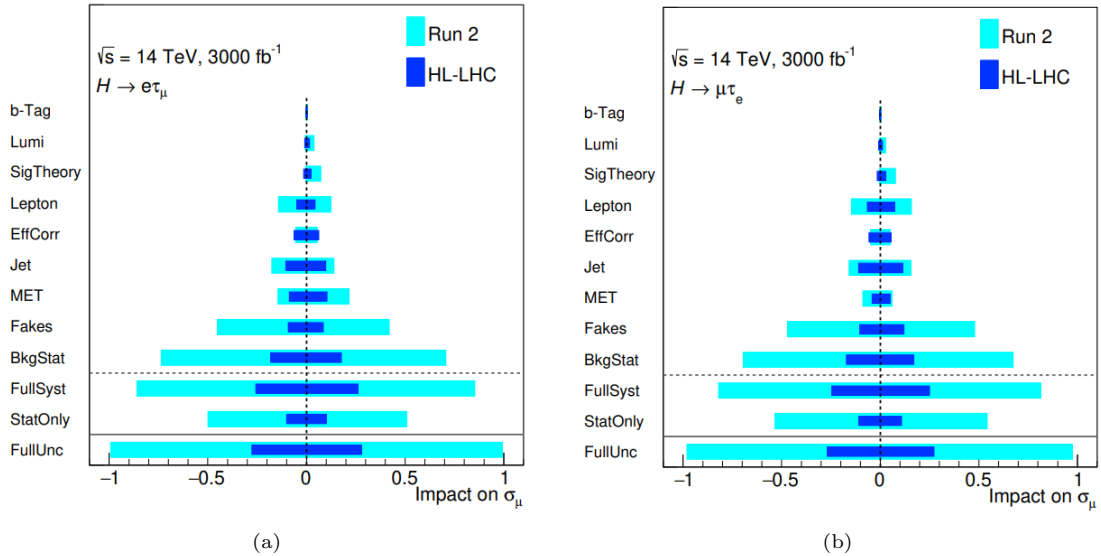


Figure 15.8: The graphical view of the impact of different types of uncertainties on the σ_μ for the (a) $e\tau$ and (b) $\mu\tau$ channels. The Run-2 analysis results are included for comparison.

Chapter 16

Conclusion and Outlook

Two direct searches for the lepton-flavour-violating decays of the Higgs boson, $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$, are presented in this thesis. They are performed with the ATLAS Run-2 dataset of proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1} .

Two background estimation methods were employed: the MC-template and symmetry methods. The MC-template method, which mainly relies on the MC simulation, is applied in the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channels. The symmetry method exploits the symmetry between electrons and muons in the SM backgrounds and is only used in the $\ell\tau_{\ell'}$ channel. The studies of the $\ell\tau_{had}$ channel using the symmetry method are still ongoing.

The Large Hadron Collider (LHC) at CERN is the world's largest and highest-energy particle collider, and ATLAS is the largest general-purpose detector at CERN. During the second long shut-down period, the ATLAS detector has been upgraded to enhance its capabilities for particle physics researches. The upgrades are focused on improving the performance of the inner tracker, energy measurements, trigger system, and data acquisition system. The Higgs LFV analysis with the Run-3 dataset could potentially benefit from the better event reconstruction, enhanced identification of the final state objects, and increasing the sensitivity of LFV signal. A short term idea could be performing the Higgs LFV analysis with partial Run-3 dataset (2022-2023) for validating the performance and the Run-2 analysis methods, while a long term perspective could be performing the full Run-3 analysis and potentially combined with the Run-2 dataset for maximizing the statistical power.

Using the BDT score distributions from the MC-template method as the final discriminant, a simultaneous 2-POI fit of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ signals was performed combining both the $\ell\tau_{\ell'}$ and $\ell\tau_{had}$ channels. The simultaneous 2-POI fit results revealed a slight excess compared to the expected SM backgrounds. However, no significant excess of events was observed to claim a new signal, both in the results of the simultaneous 2-POI fit and independent 1-POI fits. Consequently, upper limits are measured on the branching ratios of the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$, and the observed (expected) values are 0.20% (0.12%) and 0.18% (0.09%), respectively. The simultaneous 2-POI fit results indicate that the branching ratio values for both processes are consistent with zero within 2.1σ .

The results obtained using symmetry method are presented as the branching ratio difference, which is measured based on the data in $\ell\tau_{\ell'}$ final state. The best-fit value for the difference between the branching ratios of $H \rightarrow \mu\tau$ and $H \rightarrow e\tau$ ($\mathcal{B}(H \rightarrow \mu\tau) - \mathcal{B}(H \rightarrow e\tau)$) is $(0.25 \pm 0.10)\%$. This

value is compatible with zero within 2.5σ . Additionally, it is also consistent within 2σ with the best-fit value of $(0.02 \pm 0.12)\%$, obtained using the MC-template method in the $\ell\tau_{\ell'}$ final state.

By extrapolating the Run-2 analysis results, the prospect of the searches for $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ at the High-Luminosity Large Hadron Collider with the ATLAS detector have been evaluated. In the HL-LHC scenario, the integrated luminosity is assumed to achieve 3000 fb^{-1} , with the centre-of-mass energy $\sqrt{s} = 14 \text{ TeV}$. The extrapolation is done for both the MC-template and symmetry method. By performing the simultaneous 2-POI fit and independent 1-POI fits with the Asimov dataset and extrapolated samples, the expected upper limits on the branching ratio $\mathcal{B}(H \rightarrow e\tau)$ and $\mathcal{B}(H \rightarrow \mu\tau)$ obtained using MC-template method are $0.024^{+0.010}_{-0.007}\%$ for both processes. Comparing to the Run-2 analysis results, improvements with a factor of 4.8 and 3.9 are achieved for the $H \rightarrow e\tau$ and $H \rightarrow \mu\tau$ searches, respectively. The largest improvement is observed in the VBF category due to a significant reduction in the statistical uncertainties. While the symmetry method is still expected to be dominated by the statistical uncertainty from the HL-LHC data, and the expected improvements are factors of around 4.8-4.9, indicating a greater improvement compare to the MC-template method in the HL-LHC scenario.

Bibliography

- [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: [1207.7214 \[hep-ex\]](#) (cit. on pp. 1, 20).
- [2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](#) (cit. on pp. 1, 20).
- [3] Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562, arXiv: [hep-ex/9807003 \[hep-ex\]](#) (cit. on pp. 1, 23).
- [4] L. Wolfenstein, *Neutrino Oscillations in Matter*, *Phys. Rev. D* **17** (1978) 2369 (cit. on pp. 1, 23).
- [5] S. Bressler, A. Dery and A. Efrati, *Asymmetric lepton-flavor violating Higgs boson decays*, *Physical Review D* **90** (2014), URL: <https://doi.org/10.1103/PhysRevD.90.015025> (cit. on pp. 2, 88).
- [6] G. Jaeger, *The Elementary Particles of Quantum Fields*, *Entropy* **23** (2021), ISSN: 1099-4300, URL: <https://www.mdpi.com/1099-4300/23/11/1416> (cit. on p. 3).
- [7] R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 25, 822 (cit. on pp. 4, 5, 11, 14, 16, 20, 25).
- [8] R. Pasechnik and M. Šumbera, *Different Faces of Confinement*, *Universe* **7** (2021), ISSN: 2218-1997, URL: <https://www.mdpi.com/2218-1997/7/9/330> (cit. on pp. 4, 8).
- [9] J. J. T. M. F.R.S., *XL. Cathode Rays*, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44** (1897) 293, URL: <https://doi.org/10.1080/14786449708621070> (cit. on p. 5).
- [10] A. K. Halder, A. Paliathanasis and P. G. Leach, *Noether's Theorem and Symmetry*, *Symmetry* **10** (2018), ISSN: 2073-8994, URL: <https://www.mdpi.com/2073-8994/10/12/744> (cit. on p. 7).
- [11] S. Scherer and M. R. Schindler, *A Chiral Perturbation Theory Primer*, 2005, URL: <https://arxiv.org/abs/hep-ph/0505265> (cit. on p. 8).
- [12] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, *Phys. Rev.* **96** (1 1954) 191, URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191> (cit. on p. 8).
- [13] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, *Nuovo Cim.* **4** (1956) 848 (cit. on p. 8).

- [14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson,
Experimental Test of Parity Conservation in Beta Decay, **Phys. Rev.** **105** (4 1957) 1413,
URL: <https://link.aps.org/doi/10.1103/PhysRev.105.1413> (cit. on p. 9).
- [15] M. Kobayashi and T. Maskawa,
CP-Violation in the Renormalizable Theory of Weak Interaction,
Progress of Theoretical Physics **49** (1973) 652 (cit. on p. 11).
- [16] S. L. Glashow, *The renormalizability of vector meson interactions*,
Nuclear Physics **10** (1959) 107, ISSN: 0029-5582,
URL: <https://www.sciencedirect.com/science/article/pii/0029558259901968>
(cit. on p. 11).
- [17] A. Salam and J. Ward, *Electromagnetic and weak interactions*,
Physics Letters **13** (1964) 168, ISSN: 0031-9163,
URL: <https://www.sciencedirect.com/science/article/pii/0031916364907115>
(cit. on p. 11).
- [18] G. 't Hooft, *Renormalizable Lagrangians for Massive Yang-Mills Fields*,
Nucl. Phys. B **35** (1971) 167, ed. by J. C. Taylor (cit. on p. 11).
- [19] S. L. Glashow, *Partial-symmetries of weak interactions*, **Nuclear Physics** **22** (1961) 579,
ISSN: 0029-5582,
URL: <https://www.sciencedirect.com/science/article/pii/0029558261904692>
(cit. on p. 11).
- [20] J. Goldstone, *Field Theories with Superconductor Solutions*, **Nuovo Cim.** **19** (1961) 154
(cit. on p. 13).
- [21] Wikipedia, *Higgs Boson*, https://en.wikipedia.org/wiki/Higgs_boson,
Accessed: 2022-02-03 (cit. on p. 15).
- [22] L. H. W. Group, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG>,
Accessed: 2022-02-03 (cit. on pp. 16–18).
- [23] M. Schmaltz and D. Tucker-Smith, *LITTLE HIGGS THEORIES*,
Annual Review of Nuclear and Particle Science **55** (2005) (cit. on p. 21).
- [24] C. A. Baker et al.,
Improved Experimental Limit on the Electric Dipole Moment of the Neutron,
Physical Review Letters **97** (2006) (cit. on p. 22).
- [25] B. A. et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*,
Phys. Rev. Lett. **126** (14 2021) 141801 (cit. on p. 22).
- [26] D. P. Aguillard et al.,
Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm, 2023,
arXiv: [2308.06230](https://arxiv.org/abs/2308.06230) [[hep-ex](https://arxiv.org/archive/hep)] (cit. on p. 22).
- [27] M. C. et al.,
Supersymmetric explanation of the muon $g-2$ anomaly with and without stable neutralino,
Journal of High Energy Physics **2022** (2022),
URL: <https://doi.org/10.1007%2Fjhep08%282022%29124> (cit. on p. 22).

- [28] Y. Omura, E. Senaha and K. Tobe, *Lepton-flavor-violating Higgs decay $h \rightarrow \mu\tau$ and muon anomalous magnetic moment in a general two Higgs doublet model*, [Journal of High Energy Physics](#) **2015** (2015),
URL: <https://doi.org/10.1007%2Fjhep05%282015%29028> (cit. on p. 22).
- [29] W.-S. Hou and G. Kumar, *Charged lepton flavor violation in light of muon $g-2$* , [The European Physical Journal C](#) **81** (2021),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-021-09939-3> (cit. on p. 22).
- [30] S. Nie and M. Sher, *The Anomalous magnetic moment of the muon and Higgs mediated flavor changing neutral currents*, [Phys. Rev. D](#) **58** (1998), arXiv: [hep-ph/9805376](#) (cit. on p. 22).
- [31] R. Harnik, J. Kopp and J. Zupan, *Flavor violating Higgs decays*, [Journal of High Energy Physics](#) (2013),
URL: <https://doi.org/10.1007%2Fjhep03%282013%29026> (cit. on pp. 23, 25, 26, 29).
- [32] G. Blankenburg, J. Ellis and G. Isidori, *Flavour-changing decays of a 125 GeV Higgs-like particle*, [Physics Letters B](#) (2012) **5**,
URL: <https://doi.org/10.1016%2Fj.physletb.2012.05.007> (cit. on pp. 23, 26).
- [33] G. W. Bennett et al.,
Final report of the E821 muon anomalous magnetic moment measurement at BNL, [Physical Review D](#) **73** (2006), URL: <https://doi.org/10.1103%2Fphysrevd.73.072003> (cit. on p. 23).
- [34] A. Vicente, *Higgs Lepton Flavor Violating Decays in Two Higgs Doublet Models*, [Frontiers in Physics](#) **7** (2019) (cit. on pp. 23, 28–31).
- [35] T. A. et al.,
High-precision measurement of the $i\hat{z}W_i/i\hat{z}$ boson mass with the CDF II detector, [Science](#) **376** (2022) 170 (cit. on p. 23).
- [36] ATLAS Collaboration, *Improved W boson Mass Measurement using 7 TeV Proton-Proton Collisions with the ATLAS Detector*, ATLAS-CONF-2023-004, 2023,
URL: <https://cds.cern.ch/record/2853290> (cit. on p. 23).
- [37] R. Davis, D. S. Harmer and K. C. Hoffman, *Search for Neutrinos from the Sun*, [Phys. Rev. Lett.](#) **20** (21 1968) (cit. on p. 24).
- [38] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, [Zh. Eksp. Teor. Fiz.](#) **53** (1967) (cit. on p. 24).
- [39] T. M. Collaboration, *Search for the Lepton Flavour Violating Decay $\mu^+ \rightarrow e^+\gamma$ with the Full Dataset of the MEG Experiment*, 2016, arXiv: [1605.05081 \[hep-ex\]](#) (cit. on p. 25).
- [40] W.-C. Chiu, C. Geng and D. Huang, *Correlation Between Muon $g-2$ and $\mu \rightarrow e\gamma$* , [Physical Review D](#) **91** (2014) (cit. on pp. 25, 31).
- [41] U. B. et al., *Search for the decay $\mu^+ \rightarrow e^+e^+e^-$* , [Nuclear Physics B](#) **299** (1988), ISSN: 0550-3213 (cit. on p. 25).
- [42] K. H. et al., *Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs*, [Physics Letters B](#) **687** (2010) (cit. on p. 25).

- [43] B. e. a. Aubert,
Searches for Lepton Flavor Violation in the Decays $\tau^\pm \rightarrow e^\pm \gamma$ and $\tau^\pm \rightarrow \mu^\pm \gamma$,
[Phys. Rev. Lett. **104** \(2 2010\) 021802](#) (cit. on pp. 25, 171).
- [44] P. et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, en, 2013, URL: <http://cds.cern.ch/record/1559921>
(cit. on p. 25).
- [45] C. Collaboration, *Search for lepton-flavour-violating decays of the Higgs boson*,
[Physics Letters B **749** \(2015\) 337](#) (cit. on pp. 26, 27).
- [46] A. Collaboration, *Search for lepton-flavour-violating $H \rightarrow \mu\tau$ decays of the Higgs boson with the ATLAS detector*, [Journal of High Energy Physics **2015** \(2015\)](#) (cit. on pp. 26, 27).
- [47] A. Collaboration,
Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector,
[The European Physical Journal C **77** \(2017\)](#) (cit. on pp. 26, 27).
- [48] C. Collaboration, *Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and $e\tau$ in proton-proton collisions at $\sqrt{s}=13$ TeV*,
[Journal of High Energy Physics **2018** \(2018\)](#) (cit. on pp. 26, 27).
- [49] A. Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s}=13$ TeV pp collisions with the ATLAS detector*, [Physics Letters B **800** \(2020\) 135069](#)
(cit. on pp. 26, 27, 93).
- [50] C. Collaboration, *Search for lepton-flavor violating decays of the Higgs boson in the $\mu\tau$ and $e\tau$ final states in proton-proton collisions at $\sqrt{s}=13$ TeV*, [Phys. Rev. D **104** \(3 2021\) 032013](#)
(cit. on pp. 26, 27, 188).
- [51] A. Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, 2023,
[arXiv: 2302.05225 \[hep-ex\]](#)
(cit. on pp. 27, 87, 89, 96, 97, 100, 105, 107, 121, 122, 171, 177–181, 184–186, 188).
- [52] B. Grzadkowski, M. Iskrzyński, M. Misiak and J. Rosiek,
Dimension-six terms in the Standard Model Lagrangian,
[Journal of High Energy Physics **2010** \(2010\)](#) (cit. on p. 27).
- [53] J. L. Diaz-Cruz and J. J. Toscano,
Lepton flavor violating decays of Higgs bosons beyond the standard model,
[Physical Review D **62** \(2000\)](#) (cit. on pp. 27, 28).
- [54] P. S. et al., *Higgs-mass predictions in the MSSM and beyond*,
[The European Physical Journal C **81** \(2021\)](#) (cit. on p. 32).
- [55] A. Dedes, J. Ellis and M. Raidal, [Physics Letters B **549** \(2002\) 159](#) (cit. on pp. 32, 33).
- [56] A. Brignole and A. Rossi, *Lepton flavour violating decays of supersymmetric Higgs bosons*,
[Physics Letters B **566** \(2003\)](#), ISSN: 0370-2693 (cit. on p. 33).
- [57] K. Agashe, R. Contino and A. Pomarol, *The minimal composite Higgs model*,
[Nuclear Physics B **719** \(2005\)](#) (cit. on p. 33).

- [58] A. Azatov, M. Toharia and L. Zhu,
Higgs mediated flavor changing neutral currents in warped extra dimensions,
[Physical Review D](#) **80** (2009) (cit. on p. 33).
- [59] K. Agashe and R. Contino, *Composite Higgs-mediated flavor-changing neutral current*,
[Physical Review D](#) **80** (2009) (cit. on pp. 33, 34).
- [60] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*,
[Physical Review Letters](#) **83** (1999) (cit. on p. 35).
- [61] A. Azatov, M. Toharia and L. Zhu,
Higgs mediated flavor changing neutral currents in warped extra dimensions,
[Physical Review D](#) **80** (2009) (cit. on p. 35).
- [62] E. Arganda, M. J. Herrero, X. Marcano and C. Weiland, *Imprints of massive inverse seesaw model neutrinos in lepton flavor violating Higgs boson decays*, [Phys. Rev. D](#) **91** (1 2015)
(cit. on pp. 36, 37).
- [63] E. Arganda, A. M. Curiel, M. J. Herrero and D. Temes,
Lepton flavor violating Higgs boson decays from massive seesaw neutrinos,
[Physical Review D](#) **71** (2005) (cit. on p. 37).
- [64] Brüning Oliver Sim et al, *LHC Design Report*, CERN-2004-003-V-1, 2004,
URL: <http://cds.cern.ch/record/782076> (cit. on pp. 38–40).
- [65] A. Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, 2022, arXiv: [2212.09379 \[hep-ex\]](#) (cit. on p. 39).
- [66] Jean-Luc Caron, *LHC Layout*, LHC-PHO-1997-060, 2020,
URL: <https://cds.cern.ch/record/841573> (cit. on p. 40).
- [67] Lopienska, Ewa, *The CERN accelerator complex, layout in 2022*,
CERN-GRAPHICS-2022-001, 2022, URL: <https://cds.cern.ch/record/2800984>
(cit. on p. 41).
- [68] CERN, *The Proton Synchrotron*,
<https://home.cern/science/accelerators/proton-synchrotron>, Accessed: 2022-04-25
(cit. on p. 42).
- [69] CERN, *The Super Proton Synchrotron*,
<https://home.cern/science/accelerators/super-proton-synchrotron>,
Accessed: 2022-04-25 (cit. on p. 42).
- [70] T. A. Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
[Journal of Instrumentation](#) **3** (2008) (cit. on pp. 42–44, 47, 49–51).
- [71] Joao Pequeno, Paul Schaffner,
How ATLAS detects particles: diagram of particle paths in the detector, CERN-EX-1301009,
2013, URL: <https://cds.cern.ch/record/1505342> (cit. on p. 45).
- [72] Joao Pequeno, *Computer generated image of the ATLAS inner detector*,
CERN-GE-0803014, 2008, URL: <https://cds.cern.ch/record/1095926> (cit. on p. 46).
- [73] Joao Pequeno, *Computer Generated image of the ATLAS calorimeter*, CERN-GE-0803015,
2008, URL: <https://cds.cern.ch/record/1095927> (cit. on p. 48).

- [74] Joao Pequeno, *Computer generated image of the ATLAS Muons subsystem*, CERN-GE-0803017, 2008, URL: <https://cds.cern.ch/record/1095929> (cit. on p. 51).
- [75] CERN, *Muon Spectrometer*, <https://atlas.cern/Discover/Detector/Muon-Spectrometer>, Accessed: 2022-04-25 (cit. on p. 50).
- [76] L. Pontecorvo, *The ATLAS muon spectrometer*, [Eur. Phys. J. C **34** \(2004\) S117](#), ed. by J. Schuhkraft et al. (cit. on p. 52).
- [77] A. Canesse, *Small-Strip Thin Gap Chambers for the Muon Spectrometer Upgrade of the ATLAS Experiment*, [PoS LHCP2020 \(2021\) 245](#), ed. by B. Mansoulie, G. Marchiori, R. Salern and T. Bos (cit. on p. 53).
- [78] ATLAS Collaboration, *ATLAS Forward Detectors*, ATLAS-PHOTO-2018-018, 2018, URL: <https://cds.cern.ch/record/2627582> (cit. on p. 54).
- [79] ATLAS Collaboration, *Proton tagging with the one arm AFP detector*, ATL-PHYS-PUB-2017-012, 2017, URL: <https://cds.cern.ch/record/2273274> (cit. on p. 55).
- [80] T. A. collaboration, *Operation of the ATLAS trigger system in Run 2*, [Journal of Instrumentation **15** \(2020\)](#), URL: <https://doi.org/10.1088%2F1748-0221%2F15%2F10%2Fp10004> (cit. on p. 55).
- [81] CERN, *Trigger and Data Acquisition*, <https://atlas.cern/Discover/Detector/Trigger-DAQ>, Accessed: 2022-04-25 (cit. on pp. 55, 56).
- [82] T. A. Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, [Journal of Instrumentation **15** \(2020\)](#), URL: <https://doi.org/10.1088%2F1748-0221%2F15%2F04%2Fp04003> (cit. on pp. 59, 60).
- [83] T. A. Collaboration, *The ATLAS Simulation Infrastructure*, [The European Physical Journal C **70** \(2010\) 823](#), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-010-1429-9> (cit. on pp. 61, 64, 65).
- [84] S. Agostinelli et al., *Geant4—a simulation toolkit*, [Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** \(2003\) 250](#), ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 61).
- [85] A. Collaboration, *Athena*, version 22.0.1, 2019, URL: <https://doi.org/10.5281/zenodo.2641997> (cit. on p. 61).
- [86] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, 2017, URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32> (cit. on p. 61).

- [87] H. Hartanto, B. Jäger, L. Reina and D. Wackerroth,
Higgs boson production in association with top quarks in the POWHEG BOX,
[Physical Review D](#) **91** (2015), URL: <https://doi.org/10.1103/2Fphysrevd.91.094003>
(cit. on p. 61).
- [88] K. Hamilton, P. Nason, E. Re and G. Zanderighi,
NNLOPS simulation of Higgs boson production, [Journal of High Energy Physics](#) **2013** (2013),
URL: <https://doi.org/10.1007/2Fjhep10%282013%29222> (cit. on p. 61).
- [89] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*,
[Computer Physics Communications](#) **191** (2015) 159,
URL: <https://doi.org/10.1016/2Fj.cpc.2015.01.024> (cit. on p. 61).
- [90] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*,
[Journal of Physics G: Nuclear and Particle Physics](#) **43** (2016) 023001,
URL: <https://doi.org/10.1088/2F0954-3899/2F43%2F2%2F023001> (cit. on p. 61).
- [91] J. Pumplin et al.,
New Generation of Parton Distributions with Uncertainties from Global QCD Analysis,
[Journal of High Energy Physics](#) **2002** (2002) 012,
URL: <https://doi.org/10.1088/2F1126-6708/2F2002%2F07%2F012> (cit. on p. 61).
- [92] T. A. Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}=7$ TeV with the ATLAS detector*,
[Journal of High Energy Physics](#) **2014** (2014),
URL: <https://doi.org/10.1007/2Fjhep09%282014%29145> (cit. on p. 61).
- [93] R. D. B. et al., *Parton distributions for the LHC run II*,
[Journal of High Energy Physics](#) **2015** (2015),
URL: <https://doi.org/10.1007/2Fjhep04%282015%29040> (cit. on p. 62).
- [94] *ATLAS Pythia 8 tunes to 7 TeV data*, (2014),
URL: <https://cds.cern.ch/record/1966419> (cit. on p. 62).
- [95] R. D. B. et al., *Parton distributions with LHC data*, [Nuclear Physics B](#) **867** (2013) 244,
URL: <https://doi.org/10.1016/2Fj.nuclphysb.2012.10.003> (cit. on p. 62).
- [96] E. Bothmann et al., *Event generation with Sherpa 2.2*, [SciPost Physics](#) **7** (2019),
URL: <https://doi.org/10.21468/2Fscipostphys.7.3.034> (cit. on p. 62).
- [97] S. Schumann and F. Krauss,
A parton shower algorithm based on Catani-Seymour dipole factorisation,
[Journal of High Energy Physics](#) **2008** (2008) 038,
URL: <https://doi.org/10.1088/2F1126-6708/2F2008%2F03%2F038> (cit. on p. 62).
- [98] H.-L. Lai et al., *New parton distributions for collider physics*, [Physical Review D](#) **82** (2010),
URL: <https://doi.org/10.1103/2Fphysrevd.82.074024> (cit. on p. 62).
- [99] D. J. Lange, *The EvtGen particle decay simulation package*,
[Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **462** (2001) 152, BEAUTY2000,
Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002,
URL: <https://www.sciencedirect.com/science/article/pii/S0168900201000894>
(cit. on p. 62).

- [100] S. Jadach, Z. Was, R. Decker and J. Kühn, *The τ decay library TAUOLA, version 2.4*, *Computer Physics Communications* **76** (1993) 361, ISSN: 0010-4655,
URL: <https://www.sciencedirect.com/science/article/pii/001046559390061G>
(cit. on p. 62).
- [101] S. Catani, D. de Florian, G. Ferrera and M. Grazzini, *Vector boson production at hadron colliders: transverse-momentum resummation and leptonic decay*, *Journal of High Energy Physics* **2015** (2015) 1,
URL: <https://doi.org/10.1007%2Fjhep12%282015%29047> (cit. on p. 63).
- [102] N. C. Benekos, G. Dedes, L. Kashif, D. Rebutti and M. Schott, *ATLAS muon spectrometer simulation and its validation algorithms*, *Journal of Physics: Conference Series* **119** (2008) 032009,
URL: <https://dx.doi.org/10.1088/1742-6596/119/3/032009> (cit. on p. 65).
- [103] D. Rebutti, K. A. Assamagan, A. Di Simone, Y. Hasegawa and N. Van Eldik, *Geant4 Muon Digitization in the ATHENA Framework*, (2007) (cit. on p. 68).
- [104] G. A. Stewart et al., *Multi-threaded software framework development for the ATLAS experiment*, *Journal of Physics: Conference Series* **762** (2016) 012024,
URL: <https://dx.doi.org/10.1088/1742-6596/762/1/012024> (cit. on p. 72).
- [105] TimAdye, *DCube Documentation*,
<https://twiki.cern.ch/twiki/bin/view/AtlasComputing/DCube>, Accessed: 2023-06-02
(cit. on p. 73).
- [106] ‘Kolmogorov–Smirnov Test’, *The Concise Encyclopedia of Statistics*,
New York, NY: Springer New York, 2008 283, ISBN: 978-0-387-32833-1,
URL: https://doi.org/10.1007/978-0-387-32833-1_214 (cit. on p. 74).
- [107] K. Pearson, ‘On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it Can be Reasonably Supposed to have Arisen from Random Sampling’,
Breakthroughs in Statistics: Methodology and Distribution, ed. by S. Kotz and N. L. Johnson,
New York, NY: Springer New York, 1992 11, ISBN: 978-1-4612-4380-9,
URL: https://doi.org/10.1007/978-1-4612-4380-9_2 (cit. on p. 74).
- [108] A. T. CP, *ATLAS Track Reconstruction – General Overview*,
<https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview>,
Accessed: 2023-05-02 (cit. on p. 77).
- [109] T. A. Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *The European Physical Journal C* **77** (2017),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-5225-7> (cit. on p. 76).
- [110] *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, (2021)
(cit. on p. 77).
- [111] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, *JINST* **11** (2016) C12060, ed. by C. Gemme and L. Rossi (cit. on p. 77).
- [112] *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, (2019)
(cit. on p. 78).

- [113] *PV Finder for ATLAS: A Deep Learning Approach to Primary Vertex Reconstruction*, (2022), URL: <https://cds.cern.ch/record/2835256> (cit. on p. 78).
- [114] T. A. Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}=13$ TeV*, *The European Physical Journal C* **79** (2019), URL: <https://doi.org/10.1140/2Fepjc%2Fs10052-019-7140-6> (cit. on pp. 78, 79, 102).
- [115] T. A. Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, *Journal of Instrumentation* **14** (2019) P12006, URL: <https://doi.org/10.1088%2F1748-0221%2F14%2F12%2FP12006> (cit. on pp. 79, 80).
- [116] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, (2008) (cit. on p. 79).
- [117] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, (2012) (cit. on p. 79).
- [118] T. A. Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *The European Physical Journal C* **77** (2017), URL: <https://doi.org/10.1140/2Fepjc%2Fs10052-017-5004-5> (cit. on p. 79).
- [119] T. A. Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data*, *Journal of Instrumentation* **14** (2019) P03017, URL: <https://doi.org/10.1088%2F1748-0221%2F14%2F03%2FP03017> (cit. on p. 79).
- [120] G. Aad et al., *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 80–82, 102).
- [121] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063, URL: <https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063> (cit. on p. 82).
- [122] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, (2015) (cit. on p. 82).
- [123] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, (2017) (cit. on p. 82).
- [124] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, (2019) (cit. on pp. 82, 126).
- [125] *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, (2022) (cit. on pp. 83, 84).
- [126] *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, (2019) (cit. on p. 83).
- [127] T. A. Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, *The European Physical Journal C* **76** (2016), URL: <https://doi.org/10.1140/2Fepjc%2Fs10052-016-4110-0> (cit. on p. 84).

- [128] T. A. Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
[The European Physical Journal C](#) **77** (2017),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-5031-2> (cit. on p. 84).
- [129] T. A. Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
[Physical Review D](#) **96** (2017), URL: <https://doi.org/10.1103%2Fphysrevd.96.072002>
(cit. on p. 84).
- [130] *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*, (2015) (cit. on p. 84).
- [131] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*, (2015)
(cit. on p. 84).
- [132] T. A. Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, [The European Physical Journal C](#) **76** (2016),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-016-4395-z> (cit. on p. 84).
- [133] T. A. Collaboration,
Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector,
[The European Physical Journal C](#) **77** (2017),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-5081-5> (cit. on p. 84).
- [134] A. Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, 2022, arXiv: [2211.16345 \[physics.data-an\]](#) (cit. on p. 85).
- [135] T. A. Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV.*,
[The European Physical Journal C](#) **78** (2018),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-018-6288-9> (cit. on p. 86).
- [136] T. A. Collaboration,
Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector,
[Journal of High Energy Physics](#) **2015** (2015) (cit. on p. 88).
- [137] J. H. Friedman, *Stochastic gradient boosting*, [Computational Statistics and Data Analysis](#) **38** (2002) 367, Nonlinear Methods and Data Mining, ISSN: 0167-9473,
URL: <https://www.sciencedirect.com/science/article/pii/S0167947301000652>
(cit. on p. 88).
- [138] A. Elagin, P. Murat, A. Pranko and A. Safonov,
A new mass reconstruction technique for resonances decaying to $\tau^+ \tau^-$,
[Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment](#) **654** (2011) 481, ISSN: 0168-9002,
URL: <https://www.sciencedirect.com/science/article/pii/S0168900211014112>
(cit. on p. 90).
- [139] T. A. Collaboration,
Performance of electron and photon triggers in ATLAS during LHC Run 2,
[The European Physical Journal C](#) **80** (2020),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-019-7500-2> (cit. on p. 93).

- [140] T. A. Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *Journal of Instrumentation* **15** (2020) P09015, URL: <https://doi.org/10.1088%2F1748-0221%2F15%2F09%2FP09015> (cit. on p. 93).
- [141] T. A. Collaboration, *Performance of the ATLAS trigger system in 2015*, *The European Physical Journal C* **77** (2017), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-4852-3> (cit. on p. 102).
- [142] T. A. Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *The European Physical Journal C* **80** (2020), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-019-7500-2> (cit. on p. 102).
- [143] T. A. Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *Journal of Instrumentation* **15** (2020) P09015, URL: <https://doi.org/10.1088%2F1748-0221%2F15%2F09%2FP09015> (cit. on p. 102).
- [144] J. Li and H. Sun, *An Attention Based Neural Network for Jet Tagging*, 2020, arXiv: 2009.00170 [hep-ph] (cit. on p. 126).
- [145] J. A. Martinez, O. Cerri, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Pileup mitigation at the Large Hadron Collider with Graph Neural Networks*, 2019, arXiv: 1810.07988 [hep-ph] (cit. on p. 126).
- [146] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2009, arXiv: physics/0703039 [physics.data-an] (cit. on pp. 129, 131, 132, 134, 136, 137).
- [147] A. Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *Journal of Instrumentation* **14** (2019) P12006, URL: <https://doi.org/10.1088%2F1748-0221%2F14%2F12%2FP12006> (cit. on p. 142).
- [148] A. Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C* **76** (2016), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-016-4120-y> (cit. on p. 143).
- [149] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, (2017) (cit. on p. 143).
- [150] A. Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *The European Physical Journal C* **77** (2017), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-5031-2> (cit. on p. 144).
- [151] A. Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ GeV recorded in 2010 with the ATLAS detector*, *The European Physical Journal C* **73** (2013), URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-013-2306-0> (cit. on p. 144).
- [152] *E_{miss}^T performance in the ATLAS detector using 2015-2016 LHC pp collisions*, (2018) (cit. on p. 145).
- [153] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2019) (cit. on p. 145).

- [154] G. Avoni et al.,
The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS,
Journal of Instrumentation **13** (2018) P07017,
URL: <https://dx.doi.org/10.1088/1748-0221/13/07/P07017> (cit. on p. 145).
- [155] J. K. Anders and M. D’Onofrio,
V+Jets theoretical uncertainties estimation via a parameterisation method, tech. rep.,
CERN, 2016, URL: <https://cds.cern.ch/record/2125718> (cit. on p. 149).
- [156] L. Lönnblad and S. Prestel, *Matching tree-level matrix elements with interleaved showers*,
Journal of High Energy Physics **2012** (2012),
URL: <https://doi.org/10.1007%2Fjhep03%282012%29019> (cit. on p. 149).
- [157] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne,
Parton distributions in the LHC era: MMHT 2014 PDFs,
The European Physical Journal C **75** (2015),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-015-3397-6> (cit. on p. 150).
- [158] S. Dulat et al.,
New parton distribution functions from a global analysis of quantum chromodynamics,
Physical Review D **93** (2016), URL: <https://doi.org/10.1103%2Fphysrevd.93.033006>
(cit. on p. 150).
- [159] M. Bähr et al., *Herwig⁺⁺ physics and manual*,
The European Physical Journal C **58** (2008) 639,
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-008-0798-9> (cit. on p. 150).
- [160] J. Bellm et al., *Herwig 7.0/Herwig⁺⁺ 3.0 release note*,
The European Physical Journal C **76** (2016),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-016-4018-8> (cit. on p. 150).
- [161] *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*, tech. rep.,
CERN, 2017, URL: <https://cds.cern.ch/record/2261938> (cit. on p. 150).
- [162] S Dawson et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, en, 2011,
URL: <http://cds.cern.ch/record/1318996> (cit. on p. 150).
- [163] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, en, 2017,
URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32> (cit. on p. 151).
- [164] C. T. e. a. Potter, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, en, 2013,
URL: <http://cds.cern.ch/record/1559921> (cit. on p. 151).
- [165] A. H. et al., *Documentation for TRexFitter*,
<https://trexfitter-docs.web.cern.ch/trexfitter-docs/> (cit. on pp. 151, 163).
- [166] C. et al, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*,
tech. rep., New York U., 2012, URL: <https://cds.cern.ch/record/1456844>
(cit. on p. 153).
- [167] *Statistical Methods for Data Analysis*, Springer Cham, 2023 (cit. on p. 153).

- [168] J. Neyman,
Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability,
[Phil. Trans. Roy. Soc. Lond. A](#) **236** (1937) 333 (cit. on p. 158).
- [169] G. Cowan, K. Cranmer, E. Gross and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics,
[The European Physical Journal C](#) **71** (2011),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0> (cit. on pp. 159–163).
- [170] A. L. Read, *Presentation of search results: the CLs technique*,
[Journal of Physics G: Nuclear and Particle Physics](#) **28** (2002),
URL: <https://dx.doi.org/10.1088/0954-3899/28/10/313> (cit. on pp. 162, 192).
- [171] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003,
arXiv: [physics/0306116 \[physics.data-an\]](#) (cit. on p. 163).
- [172] L. Moneta et al., *The RooStats Project*, 2011, arXiv: [1009.1003 \[physics.data-an\]](#)
(cit. on p. 163).
- [173] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, HEPData (collection),
<https://doi.org/10.17182/hepdata.135719>, 2023 (cit. on pp. 171, 192, 193).
- [174] A. H. et al., *Evaluating Compatibility*, https://trexfitter-docs.web.cern.ch/trexfitter-docs/model_studies/compatibility/ (cit. on p. 186).
- [175] M. Testa and S. Jezequel, *HiggsProspects*,
<https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HiggsProspects>
(cit. on p. 189).
- [176] A. U. Physics, *HighLumiLhcSystematics2018*, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HighLumiLhcSystematics2018>
(cit. on p. 189).
- [177] R. A. Khalek, S. Bailey, J. Gao, L. Harland-Lang and J. Rojo,
Towards ultimate parton distributions at the high-luminosity LHC,
[The European Physical Journal C](#) **78** (2018),
URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-018-6448-y> (cit. on p. 191).
- [178] *Searches of lepton-flavour-violating decays of the Higgs bosons with the ATLAS detector at the HL-LHC*, tech. rep., All figures including auxiliary figures are available at
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-054>: CERN, 2022, URL: <http://cds.cern.ch/record/2841245>
(cit. on pp. 192–194, 198–200).